Title
Assessing gene effects on the brain and risk for disease using machine learning

Author
Kohannim, Omid

Publication Date
2012

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
Los Angeles

Assessing gene effects on the brain and risk for disease with machine learning

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy
in Biomedical Engineering

by

Omid Kohannim

2012

ABSTRACT OF THE DISSERTATION

Assessing gene effects on the brain and risk for disease with machine learning

by

Omid Kohannim

Doctor of Philosophy in Biomedical Engineering

University of California, Los Angeles, 2012

Professor Paul M. Thompson, Chair

The advent of neuroimaging has provided an invaluable tool for investigating brain disorders with quantitative measurements. Neuroimaging-derived measures can not only serve as biomarkers to track the progression of brain disorders like Alzheimer's disease, but also provide quantitative, intermediate phenotypes or endophenotypes close to the biology of disease, which can be utilized for finding new genes in association with brain pathology. The latter has led to the creation of a new and expanding field called neuroimaging genetics. Most methodologies for the application of neuroimaging and other biomarkers to disease diagnosis and clinical trial design utilize only single biomarkers. Similarly, in neuroimaging genetics studies, genetic variants are typically considered one by one, in association with neuroimaging phenotypes. My dissertation introduces new automated, machine learning and multivariate approaches, which potentially offer more power to biomarker-based diagnosis and clinical trial design as well as discovery and risk prediction in neuroimaging genetics.

The dissertation of Omid Kohannim is approved.

Henry Huang

Marvin Bergsneider

Arthur W. Toga

Paul M. Thompson, Committee Chair

University of California, Los Angeles
2012

To Diana, for all her love and support

TABLE OF CONTENTS

Chapter 1. Introduction viii

Background 1

Overview of subjects 3

Organization of dissertation 4

References 4

Chapter 2. National Research Service Award (NRSA) F30 Proposal 6

Specific Aims 6

Research Strategy 8

References 24

Chapter 3. Multi-biomarker classification and clinical trial enrichment 30

Introduction 30

Methods 33

Results 40

Discussion 51

References 60

Chapter 4. Multilocus analysis of brain images 66

Introduction 66

Candidate Gene Studies 67

Genome-wide associations with single imaging measures *71*

Genetic Analysis of Mass Univariate Imaging Phenotypes *74*

Multivariate imaging genetics methods *76*

Conclusion *87*

References *89*

Chapter 5. Multilocus genome-wide scan with ridge regression for gene discovery 97

Introduction *97*

Methods *98*

Results *104*

Conclusion *109*

References *109*

Chapter 6. Prediction of temporal lobe volume on MRI from genotypes with elastic net . 111

Introduction *111*

Methods *112*

Results *116*

Conclusion *119*

References *120*

Chapter 7. Multi-gene prediction of white matter structure 123

Introduction *123*

Methods *125*

Results *130*

Discussion *134*

Supplementary Information *139*

References *142*

Chapter 8. Conclusions and future directions 149

ACKNOWLEDGMENTS

I would like to first give many thanks to my advisor, Dr. Paul Thompson. Dr. Thompson has provided me with tremendous guidance and mentorship throughout my graduate years. His scientific and mathematical expertise, established role in brain imaging research, involvement with unprecedented genomics projects, and affiliation with both neurology and biomedical engineering has allowed me to pursue technically innovative, multidisciplinary projects with medically relevant outcomes. I would also like to thank my other committee members, Dr. Arthur Toga, Dr. Marvin Bergsneider, and Dr. Henry Huang, whose guidance and advice have been indispensable to my research.

I am also thankful for the invaluable resources that the Laboratory of Neuro Imaging (LONI) has supplied me during my graduate years. LONI's powerful supercomputing cluster, experienced research staff and reliable technical support have given me the unique opportunity of partaking in computationally expensive research projects. I am also grateful to Diana, all my friends and family, my group's collaborators at UCLA including Dr. Liana Apostolova and Dr. Joshua Grill, our collaborators at the University of Queensland and the Queensland Institute of Medical Research as well as colleagues in our lab, including Derrek Hibar, Neda Jahanshad, Jason Stein, Xue Hua, and others.

Research for this dissertation was supported by research grants to Omid Kohannim from the UCLA Medical Scientist Training Program (MSTP), and the National Research Service Award (NRSA; National Institute of Health Grant F30 AG041681), along with the UCLA Dissertation

Year Fellowship award. Data collection and sharing for ADNI-related projects was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Abbott, AstraZeneca AB, Bayer Schering Pharma AG, Bristol-Myers Squibb, Eisai Global Clinical Development, Elan Corporation, Genentech, GE Healthcare, GlaxoSmithKline, Innogenetics, Johnson and Johnson, Eli Lilly and Co., Medpace, Inc., Merck and Co., Inc., Novartis AG, Pfizer Inc, F. Hoffman-La Roche, Schering-Plough, Synarc, Inc., as well as non-profit partners the Alzheimer's Association and Alzheimer's Drug Discovery Foundation, with participation from the U.S. Food and Drug Administration. Private sector contributions to ADNI are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles. This research was also supported by NIH grants P30 AG010129, K01 AG030514, and the Dana Foundation. Algorithm development for this study was also funded by the NIA, NIMH, NIBIB, NICHD, the National Library of Medicine, and the National Center for Research Resources (AG016570, MH089722, EB01651, LM05639, RR019771, EB008432, EB008281, and EB007813, to Paul Thompson). The Australian twins and siblings study was supported by the National Institute of Child Health and Human Development (R01 HD050735), and the National Health and Medical Research Council (NHMRC 486682), Australia. Genotyping was supported by NHMRC (389875).

Dr. Paul Thompson supervised the research covered in this dissertation. I would like to thank my co-authors for their contributions. Chapters 3 through 7 are based on the following articles:

Kohannim O, Hua X, Hibar DP, Lee S, Chou YY, Toga AW, Jack CR Jr, Weiner MW, Thompson PM, and the Alzheimer's Disease Neuroimaging Initiative. Boosting power for clinical trials using classifiers based on multiple biomarkers. Neurobiol Aging. 2010; 31(8):1429-1442.

Hibar DP*, Kohannim O*, Stein JL, Chiang MC, Thompson PM. Multilocus genetic analysis of brain images. Front Genetics. 2011; 2(73).

Kohannim O, Hibar DP, Stein JL, Jahanshad N, Jack CR Jr, Weiner MW, Toga AW, Thompson PM, and the Alzheimer's Disease Neuroimaging Initiative. Boosting Power to Detect Genetic Associations in Imaging using Multi-Locus, Genome-Wide Scans and Ridge Regression, Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on, 1855-1859. 2011.

Kohannim O, Hibar DP, Jahanshad N, Stein JL, Hua X, Toga AW, Jack CR Jr, Weiner MW, Thompson PM, and the Alzheimer's Disease Neuroimaging Initiative. Predicting Temporal Lobe Volume on MRI from Genotypes using L1-L2 Regularized Regression, Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on. 2012; in press.

Kohannim O, Jahanshad N, Braskie MN, Stein JL, Chiang MC, Reese AH, Hibar DP, Toga AW, McMahon KL, de Zubicaray GI, Medland SE, Montgomery GW, Martin NG, Wright MJ, Thompson PM. Predicting white matter integrity from multiple common genetic variants. Neuropsychopharmacology. 2012; in press.

BIOGRAPHICAL SKETCH

VITA

2005	Undergraduate Research Scholar
2006	Beckman National Scholar
2006	Vice Provost's Award for Best Review Article
2007	B.S. (Microbiology, Immunology and Molecular Genetics major, Mathematics minor), University of California, Los Angeles
2011	UCLA Genomic Analysis and Interpretation Training Grant (Declined offer)
2011	Organization for Human Brain Mapping Trainee Abstract Award
2011	UCLA Dissertation Year Fellowship
2011	Ruth L. Kirschstein National Research Service Award (NRSA) for Individual Predoctoral MD/PhD Fellows (F30)
2010-Present	Peer-Reviewer for Human Brain Mapping, Medical Image Computing and Computer-Assisted Intervention, Neuropsychopharmacology

PUBLICATIONS

Kohannim O, Jahanshad N, Braskie MN, Stein JL, Chiang MC, Reese AH, Hibar DP, Toga AW, McMahon KL, de Zubicaray GI, Medland SE, Montgomery GW, Martin NG, Wright MJ, Thompson PM. Predicting white matter integrity from multiple common genetic variants. Neuropsychopharmacology. 2012; *in press*.

Grill JD, Di L, Lu PH, Lee C, Ringman J, Apostolova LG, Chow N, *Kohannim O*, Cummings JL, Thompson PM, Elashoff D for the Alzheimer's Disease Neuroimaging Initiative. Estimating sample sizes for predementia Alzheimer's trials based on the Alzheimer's Disease Neuroimaging Initiative. Neurobiol Aging. 2012; *in press*.

Hibar DP, Jahanshad N, Stein JL, Kohannim O, Medland SE, Hansell NK, McMahon KL, de Zubicaray GI, Montgomery GW, Martin NG, Wright MJ, Toga AW, Thompson PM. Alzheimers disease risk gene, GAB2, is associated with brain volume differences in a large cohort of young, healthy twins. Twin Res Hum Genet. 2012; *in press*.

Kohannim O, Hibar DP, Jahanshad N, Stein JL, Hua X, Toga AW, Jack CR Jr, Weiner MW, Thompson PM, and the Alzheimer's Disease Neuroimaging Initiative. Predicting Temporal Lobe

Volume on MRI from Genotypes using L1-L2 Regularized Regression, Biomedical Imaging: From Nano to Macro, 2012 IEEE International Symposium on. 2012; *in press*.

Jahanshad N, *Kohannim O*, Toga AW, McMahon KL, de Zubicaray GI, Martin NG, Wright MJ, Thompson PM. Diffusion imaging protocol effects on Genetic associations, Biomedical Imaging: From Nano to Macro, 2012 IEEE International Symposium on. 2012; *in press*.

Jahanshad N, *Kohannim O,* Hibar DP, Stein JL, McMahon KL, de Zubicaray GI, Medland SE, Montgomery GW, Whitfield JB, Martin NG, Wright MJ, Toga AW, Thompson PM. Brain structure in healthy adults is related to serum transferring and the H63D polymorphism in the HFE gene. Proc Natl Acad Sci USA. 2012; 109(14):E851-E859.

Hibar DP*, *Kohannim O**, Stein JL, Chiang MC, Thompson PM. Multilocus genetic analysis of brain images. Front Genetics. 2011; 2(73). [* indicates equal authorship].

Hibar DP, Stein JL, *Kohannim O,* Jahanshad N, Saykin AJ, Shen L, Kim S, Pankratz N, Foroud T, Huentelman M, Potkin SG, Jack CR Jr, Weiner MW, Toga AW, Thompson PM and the Alzheimer's Disease Neuroimaging Initiative. Voxel-wise gene-wide association study (vGeneWAS): multivariate gene-based association testing in 731 elderly subjects. NeuroImage. 2011; 56(4):1875-1891.

Kohannim O, Hibar DP, Stein JL, Jahanshad N, Jack CR Jr, Weiner MW, Toga AW, Thompson PM, and the Alzheimer's Disease Neuroimaging Initiative. Boosting Power to Detect Genetic Associations in Imaging using Multi-Locus, Genome-Wide Scans and Ridge Regression, Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on, 1855-1859. 2011.

Hibar DP, Stein JL, *Kohannim O*, Jahanshad N, Jack CR Jr, Weiner MW, Toga AW, Thompson PM, and the Alzheimer's Disease Neuroimaging Initiative, Principal Components Regression: Multivariate, Gene-Based Tests in Imaging Genomics, Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on, 289-293. 2011.

Kohannim O, Hua X, Hibar DP, Lee S, Chou YY, Toga AW, Jack CR Jr, Weiner MW, Thompson PM, and the Alzheimer's Disease Neuroimaging Initiative. Boosting power for clinical trials using classifiers based on multiple biomarkers. Neurobiol Aging. 2010; 31(8):1429-1442.

Gargalovic PS, Arbiglin A*, *Kohannim O**, Pagnon J, Wang X, Castellani L, LeBeouf R, Peterson ML, Spear BT, Lusis AJ. Quantitative Trait Locus Mapping and Identification of Zhx2 as a Novel Regulator of Plasma Lipid Metabolism. Circ Cardiovasc Genet. 2010; 3(1):60-67. [* indicates equal authorship].

Mungrue IN, Pagnon J, *Kohannim O*, Gargalovic PS, Lusis AJ. CHAC1/MGC4504 is a novel proapoptotic component of the unfolded protein response, downstream of the ATF4-ATF3-CHOP cascade. J Immunol. 2009; 182(1):466-476.

Chapter 1. Introduction

Background

Neuropsychiatric disorders are the leading causes of disability across the world (Killewo et al., 2010). Brain imaging modalities such as magnetic resonance imaging (MRI) are becoming increasingly important for classifying diagnosis, evaluating prognosis and tracking treatment response in patients with neuropsychiatric disorders. They provide detailed, quantitative information about brain structure and function, and aspects of ongoing disease processes. Measures and maps derived from these neuroimaging modalities have been studied through automated machine learning algorithms as potential predictors of disease outcomes. These algorithms are first *trained* with a sufficiently large number of scans from subjects with known diagnostic states. The algorithms then mathematically *learn* the patterns specific to particular classes of disease and predict, with a certain accuracy, disease states of new, *testing* subjects from their scans.

Numerous studies of neuropsychiatric disorders have proposed automated diagnostic classification algorithms. In the Alzheimer's disease field, for instance, several studies have been based on MRI scans, classifying patients based on whole-brain atrophy patterns (Davatzikos, et al., 2009), tissue density differences (Vemuri, et al., 2008) and thickness of the cerebral cortex (Lerch, et al., 2008). Automated classifiers based on neuroimaging have also been applied to aphasia (Wilson, et al., 2009), autism (Ecker, et al., 2010), psychosis (Koutsouleris, et al., 2009) and other disorders. For my dissertation research, I introduce and evaluate a computerized algorithm, based on the field of machine learning, that would incorporate not only multiple brain

imaging measures, but also biochemical and genetic information (i.e., multiple *biomarkers*) to create more powerful predictions of diagnostic and prognostic outcomes. Such novel, automated, multimodal predictors, I propose, may have important applications in future clinical decision making. For my dissertation research, I also investigate the application of such a multi-modality classifier to clinical trial design. I propose to use an automated diagnostic classifier to predict those most likely to decline, thereby reducing the size needed for a trial. This sample size reduction, which is also known as *enrichment*, can potentially make clinical trials more efficient and less costly.

In addition to its application in diagnostic classification, brain imaging also offers new quantitative phenotypes that may be closer than cognitive assessments to the underlying biological mechanisms that lead to disease (Gottesman and Gould, 2003; Meyer-Lindenberg and Weinberger, 2006; de Geus et al., 2008; Hall and Smoller et al., 2010; Marenco and Radulescu, 2010). By studying the associations of genetic factors with phenotypes based on MRI and cutting edge imaging techniques such as diffusion tensor imaging (DTI; Thomason and Thompson, 2010), mechanistically meaningful genetic contributions to brain disorders can be examined. Investigating these associations has given way to an expanding field called *neuroimaging genetics*. From a statistical perspective, these studies often take one of two broad approaches: (1) univariate or (2) multivariate. Univariate approaches consider associations between one or several genetic variants and one or several brain regions (or voxels), independently. Both neuroimaging and genetic datasets, however, are very high-dimensional with complex structures. There is, therefore, a demand for multivariate methods that may offer improved statistical power by jointly considering several genetic variants, or imaging regions (or voxels) or both. This is

discussed in more detail in Chapter 4. As part of my dissertation research, I introduce multivariate algorithms for gene discovery and risk prediction into the field of neuroimaging genetics. Algorithms that consider multiple genetic variants jointly (i.e., *multilocus* methods), I propose, are likely to (1) more powerfully detect new gene effects on brain images, and (2) identify profiles of candidate genetic variants to assist prediction of an individual's brain integrity and risk for disease.

Overview of subjects

The data for studies in my dissertation (including all neuroimaging, genetic and other data) come from two existing projects with Institutional Review Board approval and National Institute of Health funding: (a) the Queensland Twin Study in Australia and (b) the multi-center North American Alzheimer's Disease Neuroimaging Initiative (ADNI) project. For the Queensland Twin Study, identical and non-identical twins of all ages, and their siblings, born in or currently living in Queensland, Australia are considered. Neuroimaging, high-throughput genotyping, and demographic data associated with a group of these subjects (aged 21-28) have been obtained and studied. For the five-year, multi-center ADNI project, a cohort of approximately 800 elderly subjects, comprising 200 controls, 200 with Alzheimer's disease, and 400 with mild cognitive impairment, are considered. Neuroimaging, including MRI and Positron Emission Tomography (PET) data, cerebrospinal fluid marker information from lumbar punctures, as well as genotyping and demographic information have been previously collected and are accessible to the public (http://www.adni-info.org).

Organization of dissertation

My dissertation is organized into eight chapters, including the introduction and conclusions. Chapter 2 encompasses the specific aims and research strategy proposed for the successfully funded National Research Service Award (NRSA) F30, which are identical to the aims and strategies in this dissertation. In Chapter 3, automated disease classification with multiple biomarkers and its application in clinical trial design is discussed. In Chapter 4, an overview of multilocus approaches in genetic analysis of brain images is provided, which serves as a transition to the chapters that deal with applying multivariate techniques to neuroimaging genetics. Chapters 5 and 6 introduce genome-wide, multilocus approaches based on penalized regression techniques to aid in gene discovery and prediction of outcomes. Finally in Chapter 7, several neuropsychiatric risk genes are combined with the aim of personalized prediction of white matter integrity measured by DTI.

References

Davatzikos C, Fan Y, Wu X, Shen D, Resnick SM. Detection of prodromal Alzheimer's disease via pattern classification of magnetic resonance imaging. Neurobiol Aging. 2008; 29(4):514-523.

de Geus E, Goldberg T, Boomsma DI, Posthuma D. Imaging the genetics of brain structure and function. Biol Psychol. 2008; 79(1):1-8.

Ecker C, Rocha-Rego V, Johnston P, Mourao-Miranda J, Marquand A, Daly EM, Brammer MJ, Murphy C, Murphy DG; MRC AIMS Consortium. Investigating the predictive value of whole-brain structural MR scans in autism: A pattern classification approach. NeuroImage. 2010; 49(1):44-56.

Killewo J, Heggenhougen K, Quah SR. Epidemiology and Demography in Public Health. San Diego, Academic Press. 2010.

Hua X, Leow AD, Lee S, Klunder AD, Toga AW, Lepore N, Chou YY, Brun C, Chiang MC, Barysheva M, Jack CR Jr, Bernstein MA, Britson PJ, Ward CP, Whitwell JL, Borowski B, Fleisher AS, Fox NC, Boyes RG, Barnes J, Harvey D, Kornak J, Schuff N, Boreta L, Alexander GE, Weiner MW, Thompson PM, Alzheimer's Disease Neuroimaging Initiative.

3D characterization of brain atrophy in Alzheimer's disease and mild cognitive impairment using tensor-based morphometry. NeuroImage. 2008; 41(1):19-34.

Landau SM, Harvey D, Madison CM, Koeppe RA, Reiman EM, Foster NL, Weiner MW, Jagust WJ; the Alzheimer's Disease Neuroimaging Initiative. Associations between cognitive, functional, and FDG-PET measures of decline in AD and MCI. Neurobiol Aging. 2011; 32(7):1207-1218.

Gottesman II, Gould TD. The endophenotype concept in psychiatry: etymology and strategic intentions. Am J Psychiatry. 2003; 160(4):636–645.

Hall MH, Smoller JW. A new role for endophenotypes in the GWAS era: functional characterization of risk variants. Harv Rev Psychiatry. 2010; 18(1):67–74.

Koutsouleris N, Meisenzahl EM, Davatzikos C, Bottlender R, Frodl T, Scheuerecker J, Schmitt G, Zetzsche T, Decker P, Reiser M, Möller HJ, Gaser C. Use of Neuroanatomical Pattern Classification to Identify Subjects in At-Risk Mental States of Psychosis and Predict Disease Transition. Arch Gen Psychiatry. 2009; 66(7):700-712.

Lerch JP, Pruessner J, Zijdenbos AP, Collins DL, Teipel SJ, Hampel H, Evans AC. Automated cortical thickness measurements from MRI can accurately separate Alzheimer's patients from normal elderly controls. Neurobiol Aging. 2008; 29(1):23-30.

Meyer-Lindenberg A, Weinberger DR. Intermediate phenotypes and genetic mechanisms of psychiatric disorders. Nat Rev Neurosci. 2006; 7(10):818-827.

Marenco S, Radulescu E (2010). Imaging genetics of structural brain connectivity and neural integrity markers. NeuroImage 53: 848-856.

Thomason ME, Thompson PM. Diffusion imaging, white matter, and psychopathology. Annu Rev Clin Psychol. 2011; 7:63-85.

Vemuri P, Gunter JL, Senjem ML, Whitwell JL, Kantarci K, Knopman DS, Boeve BF, Petersen RC, Jack CR Jr. Alzheimer's disease diagnosis in individual subjects using structural MR images: validation studies. NeuroImage. 2008; 39(3):1186-1197.

Wilson SM, Ogar JM, Laluz V, Growdon M, Jang J, Glenn S, Miller BL, Weiner MW, Gorno-Tempini ML. Automated MRI-based classification of primary progressive aphasia variants. NeuroImage. 2009; 47(4):1558-1567.

Chapter 2. National Research Service Award (NRSA) F30 Proposal

Specific Aims

A biomarkers has been defined as "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention", with applications in assigning specific disease diagnoses, assessing staging, severity, and prognosis of disease, as well as monitoring of response to interventions (e.g., in a drug trial) (Biomarkers Definitions Working Group, 2001). A variety of biomarkers are being developed and evaluated for psychiatric and neurological disorders. These include but are not limited to measures derived from structural magnetic resonance imaging (MRI), diffusion tensor imaging (DTI), functional MRI (fMRI), and molecular imaging methods such as positron emission tomography (PET) and single-photon emission computerized tomography (SPECT), neurophysiology like electroencephalography (EEG), along with biochemistry.

Different types of biomarkers can supply complementary information useful for diagnosis or prognosis. We propose to use novel, automated machine learning-based algorithms, to help integrate multimodal biomarker data for diagnostic classification and prognostic prediction, and provide a more efficient approach to clinical trial design by selecting subjects with greater potential to respond to future therapeutics.

In addition to structural and functional neuroimaging and biochemical biomarkers, there are genetic risk factors for many psychiatric disorders that have the promise of serving as biomarkers, possibly before standard diagnostic evidence of such disorders is detectable. We propose to incorporate candidate genes into a machine learning-based model, to predict a

subject's neuroimaging-derived outcome measures such as white matter integrity. Genotype and gene expression signatures from such candidate genes may also help in personalized prediction of early risk for brain disorders.

We specifically propose to:

1.1. Develop multi-biomarker classifiers based on neuroimaging, biochemistry, and genetics for diagnostic and prognostic prediction of cognitive decline

a. To what extent can classifiers integrate information from neuroimaging measures, biochemical markers, and genetic data, to improve prediction of cognitive decline, when compared to considering each biomarker alone? We expect to see statistically stronger predictions with automated, machine learning-based methods that combine biomarker data, given the complex structural, functional and metabolic components of cognitive impairment.

b. Which sets of biomarkers are more useful for making clinically relevant outcome predictions? We expect our classifiers to determine combinations of biomarkers, during varying stages of cognitive decline, that contribute most to the prediction of diagnosis and prognosis.

c. Can such automated multi-biomarker classifiers provide a more statistically powerful means to enrichment in a hypothetical clinical trial? We anticipate that our classification approach will offer a more efficient design approach for a hypothetical clinical trial through automated selection of sub-populations of subjects who are more likely to respond to treatment.

1.2. Develop multilocus approaches for gene discovery and personalized prediction of brain integrity through neuroimaging measures from DTI

a. Can genome-wide, multivariate approaches that consider multiple loci in joint association with neuroimaging measures (e.g., DTI fractional anisotropy) boost power to discover candidate genes that affect brain integrity? Standard approaches in discovery often consider association of genetic variations with imaging endophenotypes independently; we expect, however, that our new multilocus approaches, which take the interdependence between genetic variants (e.g., linkage disequilibrium, gene-gene interactions) into account, will help detect new candidate genes affecting variations in brain structural and integrity.

b. How well can a machine learning-based model predict a subject's personalized brain integrity on DTI based on genetics? We anticipate that an automated algorithm would be able to use a subject's unique genetic profile – based on candidate single nucleotide polymorphisms (SNPs) that affect white matter integrity – and predict their white matter structural integrity measures derived from DTI.

c. To what extent can new candidate genes affecting brain integrity (e.g., based on Aims 2a and 2b) help in making clinical predictions of cognitive impairment? Here, by adding new genetic information back into classifiers discussed in aim 1, we expect to see further improvements in predictive accuracies for pertinent clinical outcomes (e.g., predicting deterioration in cognition over a given time interval).

Research Strategy

A. Develop multi-biomarker classifiers based on neuroimaging, biochemistry, and genetics for diagnostic and prognostic prediction of cognitive decline (Specific Aim 1)

1. Significance

The search for biomarkers (defined in the "Specific Aims" section) in the fields of psychiatry and neurology has been extensive and on the rise, with the increasing need for more quantifiable and objective ways of assessing brain disorders (e.g., Singh and Rose, 2009). Various modalities of neuroimaging, in particular, now offer new ways to assess structural integrity, connectivity, neural processing and metabolic activity in many brain disorders. As neuroimaging measures and other candidate biomarkers are discovered for brain disorders, the complexity of the information on a patient's biomarker profile increases; all of them may need to be considered jointly in clinical decision making. There may be biomarkers that are redundant or complementary when considered together, and the utility of biomarkers may differ at different stages of the temporal progression of a brain disorder (Jack et al., 2010). Here we propose to develop and evaluate automated algorithms to apply the complex patterns of multiple neuroimaging, biochemical and even genetic data from an individual to help predict their personalized degree of advancement in the spectrum of disease.

Machine learning algorithms have been repeatedly employed for diagnostic and prognostic prediction in the field of neuroimaging. Many focus on using volumetric or map-based measures derived from structural magnetic resonance images (MRI) (Davatzikos et al., 2008; Klöppel et al., 2008; Sun et al., 2009), and functional MRI (fMRI) (Mourao-Miranda et al., 2005; Zhang et al., 2005). Image-based diagnostic classifiers have been tested for a variety of brain disorders

and conditions including first-episode psychosis (Sun et al., 2009), schizophrenia (Yoon et al., 2007), autism (Ecker et al., 2010), and neurodegenerative diseases such as Alzheimer's disease (Davatzikos et al., 2008; Klöppel et al., 2008). Most of these studies consider the *voxelwise* maps of brain images for each subject as features for the machine learning algorithm, and classify the subject into a disease state based *solely* on their scans. What we propose here, however, is the application of machine learning to a novel, multimodal approach to the prediction of disease states based not only on different modalities of imaging, but also on other presumably complementary biomarkers such as those in the cerebrospinal fluid (CSF), along with genotype information and gene expression data for risk genes in mental illness. Here, we test our approach specifically in subjects with cognitive impairment, but this method can also be generalized to mental disorders such as schizophrenia and bipolar disorder by our many collaborators who work in these research areas.

What we additionally propose is that a clinical application of such an automated, multi-biomarker clinical outcome prediction approach is its utility in clinical trial designs. Specifically, we aim to reduce the number of subjects needed for a hypothetical clinical trial, through a new, machine learning-based, multi-biomarker enrichment strategy. By enrichment, we refer to the concept of including subjects in clinical trials who are more advanced in the spectrum of the brain disorder under study and would thereby be more likely to respond to the therapy of interest. Currently, proposed clinical trial designs that employ enrichment base this subject selection on single biomarkers (such as imaging-based measurements) with a specific threshold (Chen et al., 2007; Hampel et al., 2009; Beckett et al., 2010; Lorenzi et al., 2010). A machine learning algorithm that integrates neuroimaging, genetics and other biomarkers can, instead, conduct this

selection of subjects through a personalized predicted score for each subject and thereby boost statistical power for clinical trials. We admit that it is unlikely that a clinical trial would use a complex computer algorithm for subject selection. What is more likely (and highly desirable) is that a sub-analysis could *also* be run assessing effects in people who the classifier predicts as most likely to decline. This avoids testing promising treatments in people who are not actually deteriorating, which can underestimate the potential for detecting disease modifying effects in those who may benefit the most. We have advocated this approach in Kohannim et al. (2010).

2. Approach

Participants: Subjects from the multi-center Alzheimer's Disease Neuroimaging Initiative (ADNI) project will be considered for this specific aim, for whom various biomarker data including neuroimaging-derived measures will be analyzed. This is discussed in more detail in the "Protection of Human Subjects" section and on the web (http://www.adni-info.org/). The applicant's mentor, Paul Thompson, has an active subcontract to analyze the ADNI data, but it does not cover the scope of work in this proposal (e.g., machine learning for prognosis and diagnosis), which is a new direction developed by the applicant Omid Kohannim. Clearly, the proposal benefits from the availability of many ongoing analyses of the ADNI dataset in the lab, but this effort is distinct from ongoing funded projects in the mentor's lab.

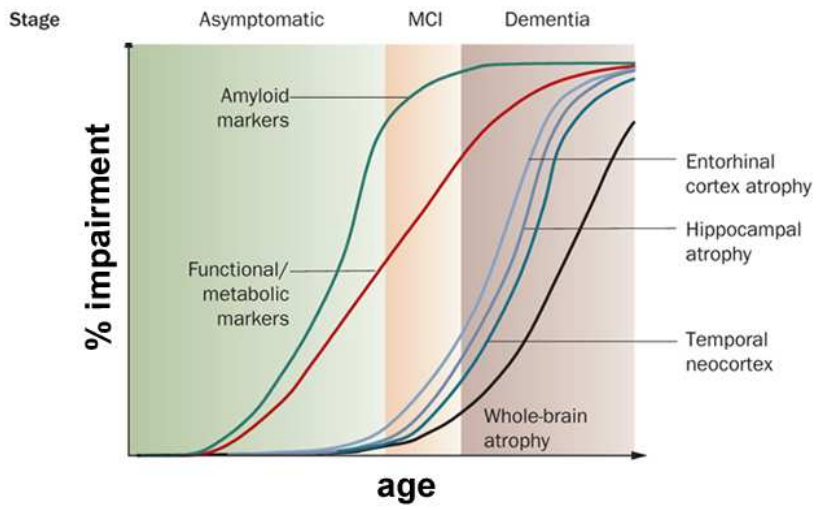
Combining Biomarkers: We will combine multiple neuroimaging-derived biomarkers in our analyses for this aim. Several summary measures will be extracted from structural MRI: (1) hippocampal volume measurements in all ADNI subjects from an automated segmentation procedure developed in our group (Morra et al., 2009), (2) ventricular volumes already acquired

for ADNI using a semi-automated method developed in our group (N=804; Chou et al., 2008), and (3) temporal lobe volumetric measures based on tensor-based morphometry (Hua et al., 2008). In addition to MRI, measurements from PET-FDG for ADNI are also available. CSF measures such as amyloid-beta, phosphorylated tau and total tau levels are also available and considered as biomarkers in ADNI. Furthermore, ADNI subjects are genotyped as described in Saykin et al. (2010) and genetic information such as their Apolipoprotein E (ApoE) genotype can also be considered as potential predictors. Notably, we do not include clinical and neuropsychological scores (e.g., from mini-mental state examination) as biomarkers for prediction of clinical outcome, as this would be circular, since the diagnoses are based on these scores. In ADNI, functional MRI, arterial spin labeling and diffusion tensor imaging information is also available for subjects, which we will additionally consider in new classifiers, if they show promise in tests we will conduct. Furthermore, improvement of diagnostic classification with the addition of gene expression data, which is available for subjects at the UCLA Easton Center, to neuroimaging, has already been noted in preliminary studies that our collaborators in Dr. Liana Apostolova's research team have published with our help (Apostolova et al., OHBM2011).

Machine Learning-Based Prediction of Clinical Outcomes: There is a wealth of machine learning algorithms that can be employed for our multi-biomarker prediction approach. These include but are not limited to discriminant analysis, artificial neural networks, decision trees and support vector machines. We decided to use the support vector machines (SVM) approach, which performs classification by maximizating a margin between diagnostic groups (Vapnik, 1999), and can also be used for regression, where continuous outcomes (e.g., clinical scores) can be predicted (Smola and Schölkopf, 2004). The main advantages of SVM are that it can handle

nonlinear problems (as many real-world relationships among variables are not linear) and unlike many other machine learning algorithms, can generalize very well to new datasets (Jain et al., 2000) - this is key for our approach to have clinical utility. In addition, SVM has been repeatedly used in the literature to classify subjects based on neuroimaging (Davatzikos et al., 2008; Ecker et al., 2010) and has demonstrated accuracy. There are different metrics to evaluate how well a classifier performs on a dataset of interest. We will compute leave-one-out accuracies for our classifiers, as this provides an unbiased evaluation (Jain et al., 2000) and is computationally feasible for our datasets. We also obtain receiver operating characteristic (ROC) curves, which are plots of true positive rates against false positive rates, and compare our classifiers' performance to random classifiers through areas under these curves. Our multimodal classification scheme will also be useful for ranking biomarkers. Various feature selection and ranking methods exist in machine learning, which we can implement to rank our biomarkers in terms of how helpful they are for a particular classification problem. One approach, which we will implement, is used for SVM classifiers in particular, where the degree to which each feature (or biomarker) contributes to the SVM weight vector determines the biomarker's ranking order (Guyon et al., 2002).

Figure 2.1. Dynamic trajectory of iomarkers (adapted from Jack et al., 2010)



We tested a pilot version of our multi-biomarker SVM approach in several subsets of the ADNI subjects, with a variety of biomarkers including neuroimaging (Kohannim et al., 2010). First, we considered the three MRI-based measures mentioned above (hippocampal, ventricular and temporal lobe volumes), along with ApoE genotypes, body mass index (BMI) and demographics (sex and age) in 737 subjects, for whom all this information was available. In a subset of 465 training subjects, leave-one-out classification accuracies of 82% and 71% were obtained for distinguishing AD from controls and mild cognitive impairment (MCI) from controls, respectively. ROC curves for the remaining 272 subjects, set aside solely for testing, displayed generalizability for the classifiers with areas under the curve of 0.95 and 0.86 respectively (**Figure 3.1**). These accuracies are comparable to studies with similar approaches (Klöppel et al., 2008; Vemuri et al., 2008) for AD, but a bit lower for MCI, which may be due to differences between the ADNI population and other cohorts, and also due to our use of summary measures from imaging instead of voxelwise maps in this analysis. Next, in a smaller subset of the ADNI subjects, we considered biochemical markers of CSF amyloid and tau, as well as a temporal lobe

summary measure from PET-FDG, in addition to MRI and ApoE. Diagnostic accuracies improved with the addition of CSF and PET-FDG separately and even further after adding both. The final multimodal classifiers showed 91% accuracy for AD and 76% accuracy for MCI in a total of 166 subjects. This demonstrates the potential to integrate biomarkers from different sources in making predictions of cognitive status (Specific Aim 1a). We also ranked the above biomarkers in their contribution to each classification, and observed that MRI-derived biomarkers were more contributory to AD classification and CSF and PET-FDG were more helpful for MCI classification. This is consistent with the recently proposed temporal sequence of AD biomarkers' dynamic trajectories (**Figure 2.1**; Jack et al., 2010) and points to a utility of our approach in providing insight to an integrative ranking system for potential biomarkers of brain disorders like cognitive impairment (Specific Aim 1b).

Power Analysis: To test the usefulness of our multi-biomarker outcome prediction approach in clinical trial designs (Specific Aim 1c), we will be conducting statistical power analyses, where the number of subjects needed to detect a clinically pertinent endpoint would be estimated for a hypothetical trial. Our goal is to reduce the sample size even further, by selecting the fraction of subjects predicted by a multi-biomarker learning algorithm to be most progressed in the spectrum of disease. This multivariate approach to enrichment, we propose, would be a promising means to making clinical trials for brain disorders more efficient, ethical and cost-effective. When we tried this approach in the ADNI datasets with SVM, as described above, we were able to obtain improvements in the number needed to detect a 25% reduction in annual temporal lobe atrophy rates with 80% power and a significance level of 0.05 (Kohannim et al.,

2010), when compared to previous literature, which compute these sample size estimates with single biomarkers like hippocampal volume (cf. Schuff et al., 2009).

B. Develop multilocus approaches for gene discovery and personalized prediction of brain integrity through neuroimaging measures from DTI (Specific Aim 2)

1. Significance

Magnetic resonance diffusion tensor imaging (DTI) was developed in the mid 1990s, and is now widely acknowledged as an invaluable tool to study white matter structure and orientation with broad applications in neuropsychiatry (Thomason and Thompson, 2011). It is sensitive to MRI signal decay due to anisotropic diffusion of water in brain tissue (Basser et al., 1994; Pierpaoli and Basser, 1996; Beaulieu et al., 2002). We will assess the standard DTI-derived measures: (1) radial and axial diffusivity, in the principal directions of the diffusion ellipsoid, which reflect the dense packing of nerve axons and their membranes, and (2) fractional anisotropy (FA), referring to the tendency for water to be constrained to diffuse in a specific direction. White matter tracts tend to constrain the diffusion of water molecules to be highly anisotropic along the tract geometries.

The implication of white matter alterations for mental illness is becoming increasingly clear at macrostructural and microstructural levels with the advent of DTI (Thomason and Thompson, 2011). DTI-based white matter abnormalities have been discovered in schizophrenia at a whole-brain and a regional level (Lim et al., 1999; Agartz et al., 2001). Anomalies in white matter integrity, particularly in prefrontal regions, have also been repeatedly observed through DTI in

bipolar disorder (Adler et al., 2004; Adler et al., 2006). Similar structural alterations in white matter have been reported in attention-deficit/hyperactivity disorder (ADHD; Ashtari et al., 2005), autism (Alexander et al., 2007) and other cognitive disorders (Fellgiebel et al., 2004).

Due to their high heritability (Chiang et al., 2011a), DTI-derived measures of white matter integrity can serve as intermediate phenotypes or "endophenotypes" (Meyer-Lindenberg and Weinberger, 2006; de Geus et al., 2008; Hall and Smoller, 2010) for powerful genome-wide searches for new risk genes in imaging genomics studies (Thompson et al., 2010). Imaging genome-wide association studies (GWAS), where imaging-derived endophenotypes are regressed against hundreds of thousands of single nucleotide polymorphisms (SNPs), such as those conducted by our group and others (Shen et al., 2010; Stein et al., 2010), are popular means to discover common genetic variants that may influence brain integrity and disease risk. With machine learning methods, SNPs can also be considered jointly in GWAS, where phenomena like epistasis and linkage disequilibrium – which render SNPs not independent of each other – may be taken into consideration with multilocus approaches (Szymczak et al., 2009). We propose the use of novel, multilocus, genome-wide searches for genes that affect brain imaging measures to supplement standard imaging GWAS for boosted detection power.

In addition to their use in gene discovery, measures from neuroimaging (e.g., voxelwise and region-of-interest FA measures from DTI) can also be used to assess genetic influences on the brain in *post hoc* and candidate gene studies. Several investigators, including our group, have already identified genetic variants, such as polymorphisms in the brain-derived neurotrophic factor (*BDNF;* Chiang et al., 2011b), clusterin (*CLU;* Braskie et al., in press), neuregulin,

catechol-*O*-methyl transferase (*COMT*), and *HFE* genes with moderate effects on white matter as detected by DTI (see below). White matter structure integrity is highly heritable (Chiang et al., 2011a) and a multitude of genetic variants are expected to have small influences on this process. Genetic polymorphisms, such as those above, can thereby contribute to a significant fraction of variability in white matter structure across individuals, which may affect their aggregate risk for developing mental illness and neurodegenerative disorders later in life. We propose that through a novel machine learning technique, we can assign personalized scores of brain integrity to individuals based on their genetic profiles. We will use our machine learning methods to integrate information from multiple genetic variations for risk assessment and to predict clinically important outcomes. In related work, gene expression profiles have been implemented in machine learning algorithms to help predict survival rates in patients with diffuse large B-cell lymphoma (Shipp et al., 2002) and that of metastasis and survival in patients with hepatocellular carcinoma (Ye et al., 2003). By incorporating a subject's genetic signature into a machine learning algorithm and predicting their DTI-derived measurements, we propose the first automated, personalized prediction study of brain integrity assessed through neuroimaging.

2. Approach

Participants: We will analyze large existing neuroimaging datasets from young adult twins as part of our collaboration with the Queensland group led by Margaret Wright in Brisbane (Thompson Genetics R01), for whom DTI scans and genotyping information is available.

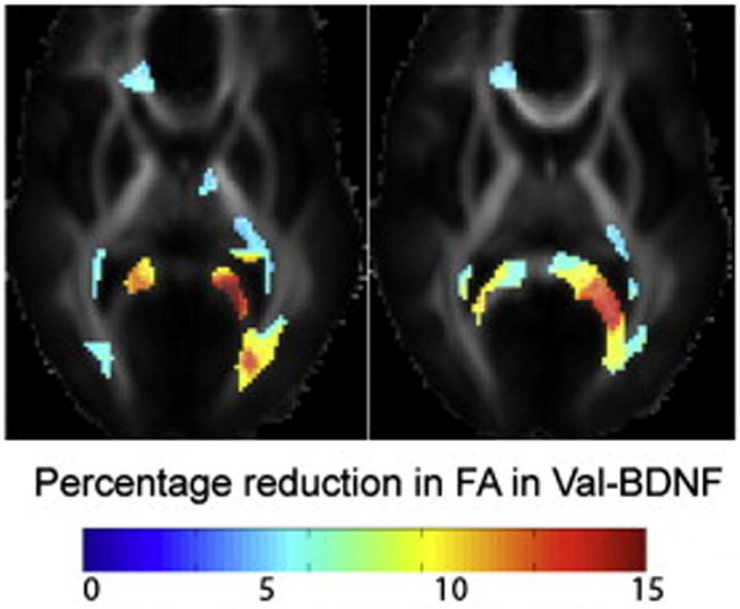
Image Acquisition and Processing: Structural and diffusion tensor whole-brain MRI scans were collected with a 4-Tesla Bruker Medspec MRI scanner. T1-weighted images were obtained with

an inversion recovery rapid gradient echo sequence. Acquisition parameters were: TI/TR/TE = 700/1500/3.35 ms; flip angle = 8°; slice thickness = 0.9 mm, with a 256×256×256 acquisition matrix. Diffusion-weighted images were also acquired using single-shot echo planar imaging with a twice-refocused spin echo sequence to reduce eddy-current induced distortions. Acquisition parameters are optimized to yield the best signal-to-noise ratio for estimation of diffusion tensors (Jones et al., 1999). Imaging parameters are: 23 cm field-of-view, TR/TE 6090/91.7 ms, with a 128 × 128 acquisition matrix. One hundred and five images are acquired for each subject: 11 with no diffusion sensitization and 94 diffusion-weighted images with gradient directions evenly distributed on the hemisphere. Standard protocols for skull-stripping, eddy correction, and adjustment for echo planar imaging artifacts and distortion as well as registration to a common coordinate space are then followed with FSL (http://www.fmrib.ox.ac.uk/fsl) tools, LONI pipeline automated tools (http://pipeline.loni.ucla.edu) and methods developed by our group and others (Leow et al., 2005; Huang et al., 2008; Jahanshad et al., 2010).

Gene Discovery: Imaging GWAS studies with DTI-based measures may lead to the discovery of new genetic risk factors that affect brain integrity. We will use our multivariate techniques for imaging GWAS, to boost statistical power to detect genetic risk factors. In the ADNI dataset, we have already found boosted associations for several SNPs in association with MRI-derived temporal lobe and hippocampal volume, with the use of a genome-wide, multilocus approach based on ridge regression (Hoerl et al., 1975; Kohannim et al., ISBI2011). This method associates neighboring SNPs jointly, in sliding windows with imaging measures, thereby adjusting for their correlation, which is due to linkage disequilibrium. We will introduce this

multilocus approach for gene discovery in DTI (Specific aim 2a), which may make possible new SNP-imaging associations, potentially undetected with standard GWAS.

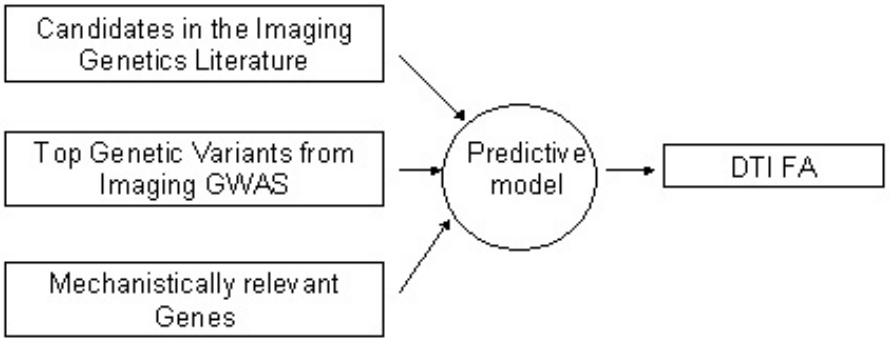
Figure 2.2. *BDNF* effects on white matter (adapted from Chiang et al., 2011b).



Candidate Genes: As mentioned above, a number of genetic variants have already been discovered with strong effects on white matter through DTI by our group and others. The well-known G196A or Val66Met polymorphism in the brain-derived neurotrophic factor (*BDNF*) gene, for instance, was recently discovered by our group to have significant modulatory effects on white matter integrity, in the splenium of the corpus callosum, inferior fronto-occipital fasciculus, left optic radiation, and corona radiata (**Figure 2.2**; Chiang et al., 2011b). This common variant in *BDNF* has been reported in many studies to increase one's risk of developing a variety of neuropsychiatric disorders (Sklar et al., 2002; Hall et al., 2003). Additionally, carriers of clusterin (*CLU*) or apolipoprotein J, which is a well-known risk factor for Alzheimer's disease (Harold et al., 2009), were recently discovered by our group to have abnormal DTI

measures (Braskie et al., in press). Our team also identified similar influences of neuregulin, which is implicated in several brain disorders such as schizophrenia (Williams et al., 2003), *HFE*, which is related to iron metabolism in the brain (Rouault and Cooperman, 2003; Kohannim et al., OHBM2011), and *FTO*, which is associated with obesity (Frayling et al., 2007) and reduced brain volume (Ho et al., 2010). Other groups find similar effects for other variants, e.g., in the dopamine-related *COMT* gene (Thomason et al., 2010). In addition to candidate gene studies, genome-wide association studies (GWAS) with DTI-derived measures (e.g., average FA values across regions of interest) and searches for mechanistically relevant genes can also lead to the discovery of more genetic variants that influence white matter architecture.

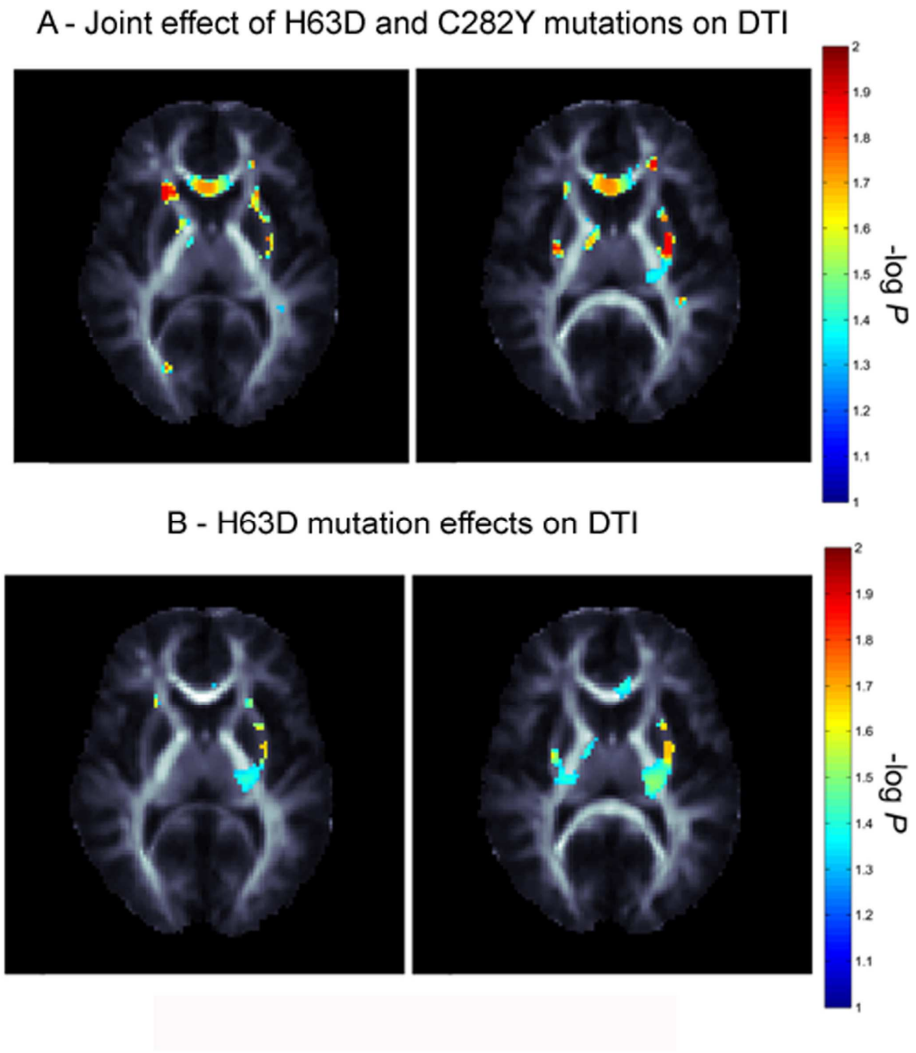
Figure 2.3. Multi-gene personalized prediction of white matter integrity



Personalized prediction of brain integrity: As discussed above, we have obtained a list of candidate genes to assess based on prior imaging genetics findings from our group, from standard and multivariate GWAS studies (summarized in Braskie et al., in press) and also from searches for genes biologically related to white matter structure and development. We will incorporate these genetic polymorphisms into a predictive model with machine learning

(Specific aim 2b; **Figure 2.3**), to predict a subject's personalized "brain integrity score" through DTI-based measurements like voxelwise FA.

Figure 2.4. Multi-SNP modeling in DTI



A first step toward such a multi-gene imaging predictor would be a standard multiple regression model, where the genetic variants are incorporated as predictors, and their joint contribution to the prediction of the DTI-based response variables can be studied. We have developed a

workflow with the use of the LONI pipeline, which would take subjects' multiple genotype information into account and regress their joint effect on the voxelwise DTI scans, yielding a *p*-value map for the combined effect of the SNPs on the subjects' images. For this pipeline, we also had to take the twins and siblings' relatedness into account. We modified and implemented the efficient mixed-model association (EMMA) code developed here at UCLA (Kang et al., 2008) to adjust for the subjects' kinship matrix for each voxelwise association. We tested this voxelwise multi-SNP model for the joint effect of two well-known, missense mutations in the *HFE* gene (H63D and C282Y) in 544 twins and siblings (Kohannim et al., OHBM2011). Significant associations were not found with C282Y alone, most likely due to its low minor allele frequency (~0.05) in our sample. Nevertheless, we discovered significant associations (satisfying multiple comparisons correction across voxels) in the genu of the corpus callosum, external capsule, superior longitudinal fasciculus and temporal lobe white matter with both C282Y and H63D (**Figure 2.4A**), with effects more extensive than those of H63D alone (**Figure 2.4B**).

We plan to use a similar multi-SNP, voxelwise computational framework for DTI in the Queensland twins and siblings dataset, but include multiple candidate genes, and employ a more powerful predictive algorithm based on machine learning (e.g., SVM regression). We will train the algorithm with the subject's genotype signature, and known FA, and test it on a new subject for a prediction of FA at a voxelwise basis, using a cross-validation scheme. Instead of *p*-values for joint associations, we will obtain correlations between predicted and known voxelwise measures, or ratios of the algorithm's prediction error divided by a "null machine's" error (a null machine simply predicts the population mean FA at each voxel).

As new genetic risk factors are discovered with DTI (Specific aim 2a) and their collective contribution to brain integrity is investigated (Specific aim 2b), we will also incorporate this new genetic information into the classifiers in aim 1. This will greatly facilitate better predictions of prognosis and at potentially earlier stages of cognitive impairment (Specific aim 2c).

References

Adler CM, Holland SK, Schmithorst V, Wilke M, Weiss KL, Pan H, Strakowski SM. Abnormal frontal white matter tracts in bipolar disorder: a diffusion tensor imaging study. Bipolar Disord. 2004; 6(3):197-203.

Adler CM, Adams J, DelBello MP, Holland SK, Schmithorst V, Levine A, Jarvis K, Strakowski SM. Evidence of white matter pathology in bipolar disorder adolescents experiencing their first episode of mania: a diffusion tensor imaging study. Am J Psychiatry. 2006; 163(2):322-324.

Agartz I, Andersson JL, Skare S. Abnormal brain white matter in schizophrenia: a diffusion tensor imaging study. Neuroreport. 2001; 12(10):2251-2254.

Alexander AL, Lee JE, Lazar M, Boudos R, DuBray MB, Oakes TR, Miller JN, Lu J, Jeong EK, McMahon WM, Bigler ED, Lainhart JE. Diffusion tensor imaging of the corpus callosum in Autism. NeuroImage. 2007; 34(1):61-73.

Apostolova L, Coppola G, Kohannim O, Thompson PM. Automated diagnostic classifiers using imaging, genotyping, and gene expression. Organization for Human Brain Mapping Conference. Quebec City, Canada. 2011.

Ashtari M, Kumra S, Bhaskar SL, Clarke T, Thaden E, Cervellione KL, Rhinewine J, Kane JM, Adesman A, Milanaik R, Maytal J, Diamond A, Szeszko P, Ardekani BA. Attention-Deficit/Hyperactivity Disorder: A Preliminary Diffusion Tensor Imaging Study. Biol Psychiatry. 2005; 57(5):448-455.

Basser PJ, Mattiello J, LeBihan D.MR Diffusion Tensor Spectroscopy and Imaging. Biophys J. 1994; 66(1):259-267.

Beaulieu C. The basis of anisotropic water diffusion in the nervous system – a technical review. NMR Biomed. 2002; 15:435-455.

Beckett LA, Harvey DJ, Gamst A, Donohue M, Kornak J, Zhang H, Kuo JH; Alzheimer's Disease Neuroimaging Initiative. The Alzheimer's Disease Neuroimaging Initiative: Annual change in biomarkers and clinical outcomes. Alzheimers Dement. 2010; 6(3):257-264.

Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther. 2001; 69(3):89-95.

Braskie MN, Jahanshad N, Stein JL, Barysheva M, McMahon KL, de Zubicaray GI, Martin NG,Wright MJ, Ringman JM, Toga AW, Thompson PM. Commonly-carried Alzheimer's disease risk gene, CLU, affects brain fiber microstructure in young adults. The Journal of Neuroscience. In press.

Braskie MN, Ringman JM, Thompson PM. Neuroimaging Measures as Endophenotypes in Alzheimer's Disease. International Journal of Alzheimer's Disease. In press.

Chen CH, Ridler K, Suckling J, Williams S, Fu CH, Merlo-Pich E, Bullmore E. Brain Imaging Correlates of Depressive Symptom Severity and Predictors of Symptom Improvement After Antidepressant Treatment. Biol Psychiatry. 2007; 62(5):407-414.

Chiang MC, McMahon KL, de Zubicaray GI, Martin NG, Hickie I, Toga AW, Wright MJ, Thompson PM. Genetics of white matter development: a DTI study of 705 twins and their siblings aged 12 to 29. NeuroImage. 2011a; 54(3):2308-2317.

Chiang MC, Barysheva M, Toga AW, Medland SE, Hansell NK, James MR, McMahon KL, de Zubicaray GI, Martin NG, Wright MJ, Thompson PM. BDNF gene effects on brain circuitry replicated in 455 twins. NeuroImage. 2011b; 55(2):448-454.

Chou YY, Leporé N, de Zubicaray GI, Carmichael OT, Becker JT, Toga AW, Thompson PM. Automated ventricular mapping with multi-atlas fluid image alignment reveals genetic effects in Alzheimer's disease. NeuroImage. 2008; 40(2):615-630.

Davatzikos C, Fan Y, Wu X, Shen D, Resnick SM. Detection of prodromal Alzheimer's disease via pattern classification of magnetic resonance imaging. Neurobiol Aging. 2008; 29(4):514-523.

de Geus E, Goldberg T, Boomsma DI, Posthuma D. Imaging the genetics of brain structure and function. Biol Psychol. 2008; 79(1):1-8.

Ecker C, Rocha-Rego V, Johnston P, Mourao-Miranda J, Marquand A, Daly EM, Brammer MJ, Murphy C, Murphy DG; MRC AIMS Consortium. Investigating the predictive value of whole-brain structural MR scans in autism: A pattern classification approach. NeuroImage. 2010; 49(1):44-56.

Fellgiebel A, Wille P, Müller MJ, Winterer G, Scheurich A, Vucurevic G, Schmidt LG, Stoeter P. Ultrastructural Hippocampal and White Matter Alterations in Mild Cognitive Impairment: A Diffusion Tensor Imaging Study. Dement Geriatr Cogn Disord. 2004; 18(1):101-108.

Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, Perry JR, Elliott KS, Lango H, Rayner NW, Shields B, Harries LW, Barrett JC, Ellard S, Groves CJ, Knight B, Patch AM, Ness AR, Ebrahim S, Lawlor DA, Ring SM, Ben-Shlomo Y, Jarvelin MR, Sovio U, Bennett AJ, Melzer D, Ferrucci L, Loos RJ, Barroso I, Wareham NJ, Karpe F, Owen KR, Cardon LR, Walker M, Hitman GA, Palmer CN, Doney AS, Morris AD, Smith GD, Hattersley AT, McCarthy MI. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science. 2007; 316(5826):889-894.

Guyon I, Weston J, Barnhill S, Vapnik V. Gene Selection for Cancer Classification using Support Vector Machines. Machine Learning. 2002. 1-3:389-422.

Hall D, Dhilla A, Charalambous A, Gogos JA, Karayiorgou M. Sequence variants of the brain-derived neurotrophic factor (BDNF) gene are strongly associated with obsessive-compulsive disorder. Am J Hum Genet. 2003; 73(2):370-376.

Hall MH, Smoller JW. A New Role for Endophenotypes in the GWAS Era: Functional Characterization of Risk Variants. Harv Rev Psychiatry. 2010; 18(1):67-74.

Hampel H, Broich K. Enrichment of MCI and early Alzheimer's disease treatment trials using neurochemical and imaging candidate biomarkers. J Nutr Health Aging. 2009; 13(4):373-375.

Harold D, Abraham R, Hollingworth P, Sims R, Gerrish A, Hamshere ML, Pahwa JS, Moskvina V, Dowzell K, Williams A, Jones N, Thomas C, Stretton A, Morgan AR, Lovestone S,

Powell J, Proitsi P, Lupton MK, Brayne C, Rubinsztein DC, Gill M, Lawlor B, Lynch A, Morgan K, Brown KS, Passmore PA, Craig D, McGuinness B, Todd S, Holmes C, Mann D, Smith AD, Love S, Kehoe PG, Hardy J, Mead S, Fox N, Rossor M, Collinge J, Maier W, Jessen F, Schürmann B, van den Bussche H, Heuser I, Kornhuber J, Wiltfang J, Dichgans M, Frölich L, Hampel H, Hüll M, Rujescu D, Goate AM, Kauwe JS, Cruchaga C, Nowotny P, Morris JC, Mayo K, Sleegers K, Bettens K, Engelborghs S, De Deyn PP, Van Broeckhoven C, Livingston G, Bass NJ, Gurling H, McQuillin A, Gwilliam R, Deloukas P, Al-Chalabi A, Shaw CE, Tsolaki M, Singleton AB, Guerreiro R, Mühleisen TW, Nöthen MM, Moebus S, Jöckel KH, Klopp N, Wichmann HE, Carrasquillo MM, Pankratz VS, Younkin SG, Holmans PA, O'Donovan M, Owen MJ, Williams J. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. Nat Genet. 2009; 41(10):1088-1093.

Ho AJ, Stein JL, Hua X, Lee S, Hibar DP, Leow AD, Dinov ID, Toga AW, Saykin AJ, Shen L, Foroud T, Pankratz N, Huentelman MJ, Craig DW, Gerber JD, Allen AN, Corneveaux JJ, Stephan DA, DeCarli CS, DeChairo BM, Potkin SG, Jack CR Jr, Weiner MW, Raji CA, Lopez OL, Becker JT, Carmichael OT, Thompson PM; Alzheimer's Disease Neuroimaging Initiative. A commonly carried allele of the obesity-related FTO gene is associated with reduced brain volume in the healthy elderly. Proc Natl Acad Sci USA. 2010; 107(18):8404-8409.

Hoerl AE, Kannard RW, Baldwin KF. Ridge regression:some simulations. Commun. Stat. Theor. Methods. 1975; 4(2):105-123.

Hua X, Leow AD, Lee S, Klunder AD, Toga AW, Lepore N, Chou YY, Brun C, Chiang MC, Barysheva M, Jack CR Jr, Bernstein MA, Britson PJ, Ward CP, Whitwell JL, Borowski B, Fleisher AS, Fox NC, Boyes RG, Barnes J, Harvey D, Kornak J, Schuff N, Boreta L, Alexander GE, Weiner MW, Thompson PM, Alzheimer's Disease Neuroimaging Initiative. 3D characterization of brain atrophy in Alzheimer's disease and mild cognitive impairment using tensor-based morphometry. NeuroImage. 2008; 41(1):19-34.

Huang H, Ceritoglu C, Li X, Qiu A, Miller MI, van Zijl PCM, Mori S. Correction of B0 susceptibility induced distortion in diffusion-weighted images using large-deformation diffeomorphic metric mapping. Magn. Reson. Imaging. 2008; 26(9):1294-1302.

Jack CR Jr, Knopman DS, Jagust WJ, Shaw LM, Aisen PS, Weiner MW, Petersen RC, Trojanowski JQ. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. Lancet Neurol. 2010; 9(1):119-128.

Jahanshad N, Lee AD, Barysheva M, McMahon KL, de Zubicaray GI, Martin NG, Wright MJ, Toga AW, Thompson PM. Genetic influences on brain asymmetry: a DTI study of 374 twins and siblings. NeuroImage. 2010; 52(2):455-469.

Jain AK, Dulin RPW, Mao J. Statistical Pattern Recognition: A Review. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2000; 22(1):4-37.

Jones DK, Horsfield MA, Simmons A. Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. Magn. Reson. Med. 1999; 42(3):515-525.

Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Eskin E. Efficient Control of Population Structure in Model Organism Association Mapping. Genetics. 2008; 178(3):1709-1723.

Klöppel S, Stonnington CM, Chu C, Draganski B, Scahill RI, Rohrer JD, Fox NC, Jack CR Jr, Ashburner J, Frackowiak RS. Automatic classification of MR scans in Alzheimer's disease. Brain. 2008; 131(Pt 3):681-689.

Kohannim O, Hua X, Hibar DP, Lee S, Chou YY, Toga AW, Jack CR Jr, Weiner MW, Thompson PM; Alzheimer's Disease Neuroimaging Initiative. Boosting power for clinical trials using classifiers based on multiple biomarkers. Neurobiol Aging. 2010; 31(8):1429-1442.

Kohannim O, Hibar DP, Stein JL, Jahanshad N, Jack CR Jr, Weiner MW, Toga AW, Thompson PM, Alzheimer's Disease Neuroimaging Initiative. Boosting Power to Detect Genetic Associations in Imaging using Multi-Locus, Genome-Wide Scans and Ridge Regression. International Symposium on Biomedical Imaging. Chicago, IL, USA. 2011.

Kohannim O, Jahanshad N, Hibar DP, Stein JL, McMahon KL, de Zubicaray GI, Medland SE, Montgomery GW, Whitfield JB, Martin NG, Wright MJ, Toga AW, Thompson PM. Investigating the Joint Effect of HFE mutations on White Matter Structure (N=544 DTI study). Organization for Human Brain Mapping Conference. Quebec City, Canada. 2011.

Leow AD, Huang SC, Geng A, Becker JT, Davis SW, Toga AW, Thompson PM. Inverse consistent mapping in 3D deformable image registration: its construction and statistical properties. IPMI. 2005; 493–503.

Lim KO, Hedehus M, Moseley M, de Crespigny A, Sullivan EV, Pfefferbaum A. Compromised White Matter Tract Integrity in Schizophrenia Inferred From Diffusion Tensor Imaging. Arch Gen Psychiatry. 1999; 56(4):367-374.

Lorenzi M, Donohue M, Paternicò D, Scarpazza C, Ostrowitzki S, Blin O, Irving E, Frisoni GB; Alzheimer's Disease Neuroimaging Initiative. Enrichment through biomarkers in clinical trials of Alzheimer's drugs in patients with mild cognitive impairment. Neurobiol Aging. 2010; 31(8):1443-1451,1451.e1.

Meyer-Lindenberg A, Weinberger DR. Intermediate phenotypes and genetic mechanisms of psychiatric disorders. Nat Rev Neurosci. 2006; 7(10):818-827.

Morra JH, Tu Z, Apostolova LG, Green AE, Avedissian C, Madsen SK, Parikshak N, Hua X, Toga AW, Jack CR Jr, Schuff N, Weiner MW, Thompson PM; Alzheimer's Disease Neuroimaging Initiative. Automated 3D Mapping of Hippocampal Atrophy and Its Clinical Correlates in 400 Subjects with Alzheimer's Disease, Mild Cognitive Impairment, and Elderly Controls. Hum Brain Mapp. 2009; 30(9):2766-2788.

Mourão-Miranda J, Bokde AL, Born C, Hampel H, Stetter M. Classifying brain states and determining the discriminating activation patterns: Support Vector Machine on functional MRI data. NeuroImage. 2005; 28(4):980-995.

Pierpaoli C, Basser PJ. Toward a Quantitative Assessment of Diffusion Anisotropy. Magn Reson Med. 1996; 36(6):893-906.

Rouault TA, Cooperman S. Brain Iron Metabolism. Semin Pediatr Neurol. 2006; 13(3):142-148.

Saykin AJ, Shen L, Foroud TM, Potkin SG, Swaminathan S, Kim S, Risacher SL, Nho K, Huentelman MJ, Craig DW, Thompson PM, Stein JL, Moore JH, Farrer LA, Green RC, Bertram L, Jack CR Jr, Weiner MW; Alzheimer's Disease Neuroimaging Initiative. Alzheimer's Disease Neuroimaging Initiative biomarkers as quantitative phenotypes: Genetics core aims, progress, and plans. Alzheimers Dement. 2010; 6(3):265-273.

Schuff N, Woerner N, Boreta L, Kornfield T, Shaw LM, Trojanowski JQ, Thompson PM, Jack CR Jr, Weiner MW; Alzheimer's Disease Neuroimaging Initiative. MRI of hippocampal volume loss in early Alzheimer's disease in relation to ApoE genotype and biomarkers. Brain. 2009; 132(Pt 4):1067-1077.

Shen L, Kim S, Risacher SL, Nho K, Swaminathan S, West JD, Foroud T, Pankratz N, Moore JH, Sloan CD, Huentelman MJ, Craig DW, Dechairo BM, Potkin SG, Jack CR Jr, Weiner

MW, Saykin AJ; Alzheimer's Disease Neuroimaging Initiative. Whole genome association study of brain-wide imaging phenotypes for identifying quantitative trait loci in MCI and AD: A study of the ADNI cohort. NeuroImage. 2010; 53(3):1051-1063.

Shipp MA, Ross KN, Tamayo P, Weng AP, Kutok JL, Aguiar RC, Gaasenbeek M, Angelo M, Reich M, Pinkus GS, Ray TS, Koval MA, Last KW, Norton A, Lister TA, Mesirov J, Neuberg DS, Lander ES, Aster JC, Golub TR. Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. Nat Med. 2002; 8(1):68-74.

Singh I, Rose N. Biomarkers in psychiatry. Nature. 2009; 460(7252):202-207.

Sklar P, Gabriel SB, McInnis MG, Bennett P, Lim YM, Tsan G, Schaffner S, Kirov G, Jones I, Owen M, Craddock N, DePaulo JR, Lander ES. Family-based association study of 76 candidate genes in bipolar disorder: BDNF is a potential risk locus. Brain-derived neutrophic factor. Mol Psychiatry. 2002; 7(6):579-593.

Smola AJ, Schölkopf B. A tutorial on support vector regression. Statistics and Computing. 2004; 14(3):199-222.

Stein JL, Hua X, Morra JH, Lee S, Hibar DP, Ho AJ, Leow AD, Toga AW, Sul JH, Kang HM, Eskin E, Saykin AJ, Shen L, Foroud T, Pankratz N, Huentelman MJ, Craig DW, Gerber JD, Allen AN, Corneveaux JJ, Stephan DA, Webster J, DeChairo BM, Potkin SG, Jack CR Jr, Weiner MW, Thompson PM; Alzheimer's Disease Neuroimaging Initiative. Genome-wide analysis reveals novel genes influencing temporal lobe structure with relevance to neurodegeneration in Alzheimer's disease. NeuroImage. 2010; 51(2):542-554.

Sun D, van Erp TG, Thompson PM, Bearden CE, Daley M, Kushan L, Hardt ME, Nuechterlein KH, Toga AW, Cannon TD. Elucidating a Magnetic Resonance Imaging-Based Neuroanatomic Biomarker for Psychosis: Classification Analysis Using Probabilistic Brain Atlas and Machine Learning Algorithms. Biol Psychiatry. 2009; 66(11):1055-1060.

Szymczak S, Biernacka JM, Cordell HJ, González-Recio O, König IR, Zhang H, Sun YV. Machine Learning in Genome-Wide Association Studies. Genet Epidemiol. 2009;33 Suppl 1:S51-7.

Thomason ME, Dougherty RF, Colich NL, Perry LM, Rykhlevskaia EI, Louro HM, Hallmayer JF, Waugh CE, Bammer R, Glover GH, Gotlib IH. COMT genotype affects prefrontal white matter pathways in children and adolescents. NeuroImage. 2010; 53(3):926-934.

Thomason ME, Thompson PM. Diffusion imaging, white matter, and psychopathology. Annu Rev Clin Psychol. 2011; 7:63-85.

Thompson PM, Martin NG, Wright MW. Imaging genomics. Curr Opin Neurol. 2010; 23(4):368-373.

Vapnik V. The Nature of Statistical Learning Theory. Springer, New York. 1995.

Vemuri P, Gunter JL, Senjem ML, Whitwell JL, Kantarci K, Knopman DS, Boeve BF, Petersen RC, Jack CR Jr. Alzheimer's disease diagnosis in individual subjects using structural MR images: validation studies. NeuroImage. 2008; 39(3):1186-1197.

Williams NM, Preece A, Spurlock G, Norton N, Williams HJ, Zammit S, O'Donovan MC, Owen MJ. Support for genetic variation in neuregulin 1 and susceptibility to schizophrenia. Mol Psychiatry. 2003; 8(5):485-487.

Ye QH, Qin LX, Forgues M, He P, Kim JW, Peng AC, Simon R, Li Y, Robles AI, Chen Y, Ma ZC, Wu ZQ, Ye SL, Liu YK, Tang ZY, Wang XW. Predicting hepatitis B virus–positive metastatic hepatocellular carcinomas using gene expression profiling and supervised machine learning. Nat Med. 2003; 9(4):416-423.

Yoon U, Lee JM, Im K, Shin YW, Cho BH, Kim IY, Kwon JS, Kim SI. Pattern classification using principal components of cortical thickness and its discriminative pattern in schizophrenia. NeuroImage. 2007; 34(4):1405-1415.

Zhang L, Samaras D, Tomasi D, Volkow N, Goldstein R. Machine Learning for Clinical Diagnosis from Functional Magnetic Resonance Imaging. Proc. IEEE Conf. Comput. Vis. Pattern Recognit. 2005; 1211-1217.

Chapter 3. Multi-biomarker classification and clinical trial enrichment

Introduction

Alzheimer's Disease (AD), the most common form of dementia, affects approximately 5.3 million people in the United States alone, and its prevalence continues to rise (Alzheimer's Association, 2009). Research and therapeutic efforts also focus on subjects with Mild Cognitive Impairment (MCI) – an intermediate condition between healthy aging and AD – as they convert to AD at a heightened rate of 10-15% per year (Petersen, et al., 1999). Multiple imaging biomarkers have been used for quantifying disease progression and measuring various aspects of AD pathology, such as amyloid and tau deposition, measured by Positron Emission Tomography (PET) and radiotracers that bind to the plaques and tangles in the brain (Klunk, et al., 2004; Protas, et al., 2010), metabolic decline or perfusion deficits assessed by fluoro-deoxyglucose PET (PET-FDG), brain atrophy on MRI, and risk factors that influence these measures (e.g. ApoE, cardiovascular risks, etc.) (Frisoni, et al., 2010; Jack, et al., 2010; Petersen, 2010).

Although the disease can be tracked in many ways, methods are also needed to integrate these multiple measures to achieve greater power in diagnosis and prognosis. Machine learning algorithms such as linear discriminant analysis, support vector machines, and boosting have recently been proposed to combine multiple AD features derived from brain imaging and other biomarkers, for AD and MCI classification. Several studies have performed diagnostic classification based on MRI scans, using measures such as whole-brain patterns of atrophy (Davatzikos, et al., 2009; Mesrob, et al., 2008), tissue densities from voxel-based morphometry (Vemuri, et al., 2008) and cortical thickness (Lerch, et al., 2008). Vemuri et al. (2008) assigned

overall "scores" for each subject's MRI – called the Structural Abnormality Index (STAND) - based on gray and white matter voxels that best differentiated AD patients from controls. In related work, Davatzikos et al. (2009) assigned "scores" to each subject's MRI scan based on a minimal set of brain regions that best discriminated AD from normal controls in a training sample; their approached is termed Spatial Pattern of Abnormality for Recognition of Early Alzheimer's disease, or SPARE-ED.

Researchers have also explored adding other predictors to improve the accuracy of MRI for computer-assisted diagnosis of AD and MCI, and for predicting whether a person will convert from MCI to AD in the near future. PET, for example, offers metabolic or perfusion-based information that complements measures of structural atrophy on MRI (Fan, et al., 2008; Hinrichs, et al., 2009). Vemuri et al. (2009) adjusted their STAND scores by incorporating demographic variables such as age, sex, and ApoE genotype, and this improved their classification accuracy. Additionally, MRI-based STAND scores were shown to improve the accuracy of CSF biomarkers for predicting cognitive decline, including total tau (t-tau), phosphorylated tau (p-tau) and the beta-amyloid isoform, $a\beta_{42}$ (Vemuri, et al., 2009).

It is worth noting that MRI-based machine learning has been used widely for classification not only for AD, but also for predicting changes in patients with brain tumors (Lukas, et al., 2004), aphasia (Wilson, et al., 2009), autism (Ecker, et al., 2010), psychosis (Koutsouleris, et al., 2009) and even for classifying patterns of brain activation in functional MRI (Mourão-Miranda, et al., 2005). Similar algorithms have been implemented to distinguish AD from other types of dementia such as frontotemporal dementia (Davatzikos, et al., 2008; Klöppel, et al., 2008).

Support vector machines (SVMs) are one of the most widely-used and effective tools for classification of AD and other neurological disorders, and are used in many of the reports listed above. We therefore set out to test how well SVMs would perform for classifying patients as having AD and MCI based on multiple imaging and biological measures in ADNI, as well as for predicting imminent cognitive decline.

A second goal of this paper was to make a conceptual connection between sample size requirements for clinical trials and the power of classifiers to predict future decline. By using our classifiers to predict those most likely to decline, we tested the hypothesis that this subset might experience atrophic rates with greater effect sizes. This concept is termed clinical trial enrichment, as it seeks out a sub-sample of subjects who might be better candidates for demonstrating therapeutic effects, at least from a statistical standpoint (see Discussion for assumptions of this approach).

We recently found that regional numerical summaries derived from tensor-based morphometry of longitudinal MRI (over a 1-year interval) can reduce the estimated sample size requirements to 48 AD and 88 MCI subjects per arm of a hypothetical clinical trial (treatment versus placebo), for detecting a 25% reduction in the mean annual temporal lobe atrophy rate with 80% power (Hua, et al., 2009). Power was similar when 3 Tesla or 1.5 Tesla MRI scans were used (Ho, et al., 2009); still higher power was possible for trials with longer follow-up intervals (Hua, et al., 2010b). Other groups report comparable power for measures based on hippocampal volumes (Schuff, et al., 2009). Through the use of multi-modality classifiers, these and other similar sample size estimates can presumably be reduced still further.

In this report, our goals were: (1) to statistically combine baseline MRI measures of hippocampal, temporal and ventricular volumes with age, sex, ApoE genotype and body mass index (BMI) for AD and MCI classification, (2) to examine how the best-performing predictors would be further enhanced by using information on CSF biomarkers and PET-FDG; (3) to evaluate this multi-modality approach for predicting cognitive decline in MCI, and, most importantly, (4) to assess whether we could expect to reduce clinical trial sample size estimates by using our classifiers to target those most likely to decline. Numerous structural MRI-based measures, including hippocampal and ventricular volumes, as well as other temporal lobe summaries, have already been validated as indicators of AD progression, particularly after the MCI stage (Frisoni, et al., 2010). We hypothesized that using multiple MRI summaries (rather than choosing one) might offer complementary information to classify patients into the correct diagnostic categories and predict cognitive decline, thereby providing a new way to boost the power of clinical trials.

Methods

A. Subjects

Baseline neuroimaging and biomarker data were downloaded from the Alzheimer's Disease Neuroimaging Initiative (ADNI) public database (http://www.loni.ucla.edu/ADNI/Data/) on or before November 20, 2009 and reflect the status of the database at that point; as data collection is ongoing. ADNI is a large five-year study launched in 2004 with the primary goal of testing whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessments at multiple sites (as in a typical clinical trial), can replicate results from smaller single site studies measuring the progression of MCI and early AD. More sensitive and specific

markers of early AD progression is will help to monitor the effectiveness of new treatments, and lessen the time and cost of clinical trials.

Available Data for Baseline Subjects. In what follows, sample sizes for analyses using different predictors are slightly different, as the study is ongoing and not all measures could be collected from all ADNI subjects. For our classification study based on baseline MRI numerical summaries, ApoE, age, sex and BMI, data were available from 737 ADNI subjects (158 AD: 75.4 ± 7.4 years of age, 366 MCI: 74.8 ± 7.3 years of age, and 213 controls: 76.0 ± 5.1 years of age). To equalize the sex distribution, we reduced the MCI subject set to a group of 264 sex-matched subjects. As there were 102 more men than women in the MCI group, we ranked the MCI males based on numbers assigned to them via a computerized random number generator and removed the first 102 to ensure that the elimination process was random and unbiased. A similar method was implemented for the other MCI groups as well. For our next classification study, we were limited by the availability of PET-FDG and CSF data, so our studies included subsets of the subjects considered above (328 subjects after adding only CSF, 364 subjects after adding only PET-FDG, and 166 subjects after adding both CSF and PET-FDG). For the first part of the cognitive decline prediction study, we considered 64 sex-matched MCI subjects, of whom 12 converted to AD in 12 months. 64 subjects remained after selecting MCI subjects who had all biomarker information and equalizing the distribution of sex. The fraction of converters here (18.75%) is a slightly higher than the previously estimated rate of conversion in ADNI (13% according to Petersen et al., 2010); the rate is marginally higher as a subgroup of male non-converters was excluded to allow sex matching. A larger sample of 129 sex matched MCI subjects with a reduced number of biomarkers was considered for the second part of the same study, 22 of whom (17.05%) converted to AD within 12 months. Sex matching was performed

through a random elimination process as described above. The subjects and biomarkers included in each study are summarized in **Table 3.1**.

Table 3.1. ADNI subjects and biomarkers included in each study. Here we outline the subject samples analyzed for different classification tests. Subjects are split into independent training and testing samples to ensure correct evaluation of classifier performance. *MRI* denotes that a 1.5T MRI scan was available; *BMI* denotes body mass index. *CSF* denotes that CSF-derived biomarkers were available.

Study	Biomarkers	Number of Subjects (Training + Testing)		
		AD	**MCI**	**CN**
1	MRI, Age, ApoE, Sex, BMI	158 (118 + 40)	264 (184 + 80)	213 (163 + 50)
2a	MRI, Age, ApoE, CSF	77 (57 + 20)	158 (118 + 40)	93 (68 + 25)
2b	MRI, Age, ApoE, PET-FDG	79 (59 + 20)	191 (146 + 45)	94 (74 + 20)
2c	MRI, Age, ApoE, CSF, PET-FDG	40 (20 + 20)	83 (43 + 40)	43 (23 + 20)
3a	MRI, Age, ApoE, CSF, PET-FDG	-	64 (41 + 23)	-
3b	MRI, ApoE, PET-FDG	-	129 (67 + 62)	-

B. Biomarkers

For each subject, the biomarkers we considered included three MRI-derived numerical summaries, a PET-FDG numerical summary, and three CSF biomarkers (t-tau, p-tau and $a\beta_{42}$), ApoE genotype (coded as 0, 1, or 2 for the number of E4 alleles), sex (coded as 0 or 1), age, and BMI. In addition to MRI, PET-FDG and CSF can provide important functional and pathological information on AD progression (Jack, et al., 2010). We also considered ApoE genotype (coded as 0, 1 or 2 for the number of E4 alleles), age, sex and BMI, as each can influence AD risk (Corder, et al., 1993, Lindsay, et al., 2002, Azad, et al., 2007, Buchman, et al., 2005). BMI was included as a number of recent studies found that higher BMI is associated with greater brain atrophy in normal elderly subjects (Raji, et al., 2009), and in MCI and AD (Ho, et al., 2010b). This effect still holds true after accounting for the effects of hypertension, diabetes, and the level of white matter hyperintensities (Ho, et al., 2010b) on the brain. In addition, a commonly carried risk gene for obesity, FTO, was recently reported to be associated with the level of brain atrophy in the ADNI cohort (Ho, et al., 2010a), so we included BMI as it is a cardiovascular risk factor associated with brain atrophy. Clinical biomarkers that were used in ADNI to determine diagnosis, such as the sum of boxes Clinical Dementia Rating (sobCDR) and other similar measures are used by physicians for making diagnoses and were therefore not used as features for classification to avoid circular inference. In fact, using sobCDR alone for classification led to almost perfect classification accuracy, as accuracy here is judged in terms of agreement with clinical diagnosis, the best available proxy when *post mortem* neuropathological data is not yet available. Instead, the annual rate of change in sobCDR was used as an outcome measure of cognitive decline to help define conversion from MCI to AD.

The MRI features included numerical summaries from the hippocampus, lateral ventricles and a TBM-derived measure of atrophy in the temporal lobes. The hippocampal summaries were volumes generated from an automatic segmentation method that we developed based on machine learning; we recently validated this method against manual gold standards (Morra, et al., 2008; Morra, et al., 2009; Morra, et al., 2010). The ventricular summaries were volumes acquired from a semi-automated, multi-atlas segmentation technique that we developed (multi-atlas fluid image alignment or *MAFIA*; (Chou, et al., 2008)). The temporal lobe summaries were obtained from an anatomically defined region-of-interest (ROI) on three-dimensional atrophy maps generated with tensor-based morphometry (Hua, et al., 2008a; Hua, et al., 2008b). PET-FDG numerical summaries were based on a pre-defined temporal lobe ROI (Landau, et al., 2009). All imaging summaries were averaged for the lobes in the left and right brain hemispheres.

CSF samples were obtained through lumbar puncture, after an overnight fast. Samples from various sites were transferred, on dry ice, to the ADNI Biomarker Core Laboratory at the University of Pennsylvania Medical Center, where the levels of t-tau, p-tau and $a\beta_{42}$ are measured with a multiplex immunoassay platform under the direction of Drs. Leslie Shaw and John Trojanowski. ApoE genotyping was performed on DNA samples from subjects' blood. Genomic DNA samples were analyzed using the Human610-Quad BeadChip (Illumina, Inc. San Diego, CA) at the University of Pennsylvania. Demographic data were obtained from **https://www.loni.ucla.edu/ADNI/Data/**. It should be emphasized that only baseline values of the biomarkers were used for prediction.

C. Support Vector Machines

SVMs are a type of machine learning or pattern recognition method that can be used to classify a dataset into different groups, based on multiple features, or measures, available for each subject (see e.g., Morra et al., 2009b). As with linear discriminant analysis, a number of observations about a subject (here the imaging and other measures) may be assembled into a vector, with as many components as there are measures. Then a mathematical function is estimated (or "learned") that best combines these features to give an output that indicates, as accurately as possible, which group the individual belongs to. For an introduction to SVM - comparing it to simpler methods such as linear discriminant analysis (LDA) - please see our tutorial (Morra et al., 2009b). As mentioned in the introduction, SVM was chosen as a machine learning algorithm for this report due to its successful performance in the previous AD literature (Davatzikos, et al., 2009, Fan, et al., 2008, Mesrob, et al., 2008, Vemuri, et al., 2008), and for other neurobiological applications (Ecker, et al., 2010, Koutsouleris, et al., 2009, Wilson, et al., 2009). SVMs may be considered as a generalization of linear regression, which use a supervised learning method to fit a classification function to the data in a training set of labeled observations. Other types of classifiers, such as adaptive boosting (Freund and Schapire, 1999; Morra et al., 2010), may also be useful for subject classification based on multiple biomarker measures, as they optimally combine predictors that individually perform weakly, but perform strongly in combination.

SVM is formulated as an optimization problem. Given a set of training data with corresponding class labels, a hyperplane is sought that maximizes the margin (a measure of the ability to differentiate) between different classes. This hyperplane, computed from a training set of example data, can then be utilized to classify newly presented (independent) testing data sets. Data consist of a set of vectors (x_1, …, x_n) where each vector contains a number of features and the class labels are scalars (y_1, …, y_n) where y_i is either 1 or -1 in a 2-class problem. The

optimization problem for a linear SVM is written as $min\ \frac{1}{2}\|w\|^2\ \ subject\ to\ \ y_i(x_i \cdot w + b) \geq 1$, where *w* and *b* represent the normal vector to and the intercept of the hyperplane respectively. For cases where a linear surface (hyperplane) cannot effectively separate the data, nonlinear kernels, such as radial basis functions (RBFs), are incorporated into the optimization problem. Additionally, "slack variables" may be introduced with a tunable parameter, *C*, to allow for a balance between misclassifications and the width of the margin. With this modification, the optimization problem may be restated as $min\ \frac{1}{2}\|w\|^2 + C\sum_i \xi_i\ subject\ to\ \ y_i(x_i \cdot w + b) \geq 1 - \xi_i$, where ξ_i is the slack variable for each *i* (Vapnik, 1998, Burges, 1998). SVMs may also be utilized for regression, where instead of a binary output, it would predict a continuous output for each subject's input vector, *x*. We performed our experiments using the LS-SVM package for classification and regression (Suykens and Vandewalle, 1999) in Matlab (MathWorks, Natick, MA).

D. Training and Testing

We divided AD, MCI and control subjects randomly into training and testing sets as shown in **Table 3.1**. The training sets were used for parameter optimization (regularization parameter *C* for a linear kernel; *C* and kernel-specific parameter, σ, for an RBF kernel) and for leave-one-out cross-validation. The SVM models were tested on independent testing sets to ensure generalizability. Receiver operating characteristic (ROC) curves were obtained to demonstrate the trade-off between sensitivity and specificity. ROC curves were compared, to evaluate different classifiers, using a statistical method developed for ROC analysis (Hanley and McNeil, 1983) in the MedCalc Statistical Software (MedCalc, Mariakerke, Belgium). When SVM was

implemented for prediction instead of classification, mean squared errors were used for comparison, instead of misclassification errors.

E. Power Analysis

A power analysis was defined by the ADNI Biostatistics Core to estimate the sample size required to detect a 25% reduction in the mean annual rate of atrophy, using a two-sided test and standard significance level (α=0.05) for a hypothetical two-arm study (treatment versus placebo), with 80% power (this number is referred to as n80, and smaller numbers are better). The formula is $n = \frac{2\sigma^2 (z_{1-\alpha/2} + z_{power})^2}{(0.25\beta)^2}$, where σ and β refer to the mean and standard deviation in the atrophic rates respectively, α is set to be 0.05, and the desired power is 80%. Atrophic rates were determined based on a statistically-defined ROI by training on 22 AD subjects, as described more fully in (Hua, et al., 2009). Brain atrophy rates measured by MRI correlate with the progression of Alzheimer's disease, and offer baseline and transitional predictive power for diagnosis, making them clinically relevant endpoints for power analysis (Duara, et al., 2008, Fox, et al., 2000, Jack, et al., 2004).

<u>*Results*</u>

A. AD and MCI Classification based on MRI markers, ApoE genotype and demographic information

We first used the 3 MRI-derived summaries, ApoE genotype and demographic variables (age, sex and BMI) for AD and MCI classification with 635 ADNI subjects. SVM training was performed with all seven features using a linear kernel with C = 1, and the contributions of the

different biomarkers were put into a rank order (best to worst) based on their SVM weights, assessed by w_i^2 in the notation of SVM described in the methods. The rank orders are shown in **Table 3.2**.

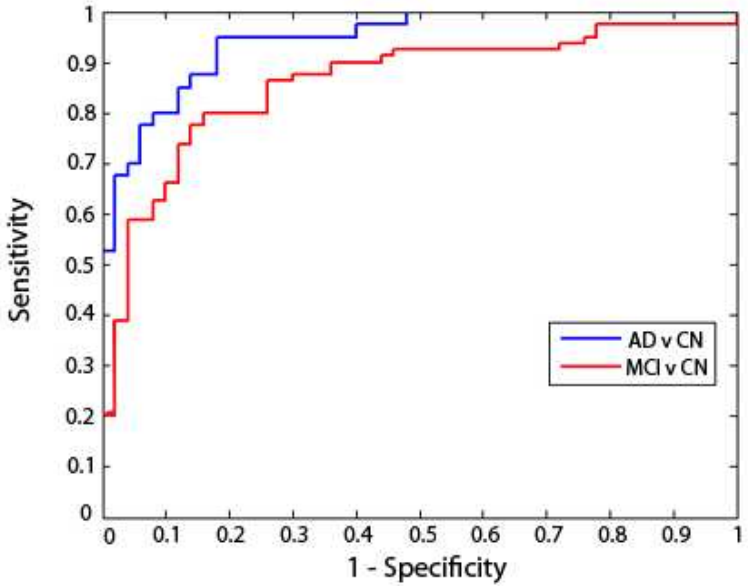
Table 3.2. Rank order list with relative SVM weights for MRI, ApoE, Age, Sex and BMI in AD and MCI classification. Hippocampal volumes were the most influential feature for differentiating AD from controls, closely followed by ApoE genotype, which outperformed all the other MRI-derived markers. For classifying subjects as either MCI or controls, the exact same features were useful, in the same order of priority. This is somewhat in line with expectation, as hippocampal volume is so widely used and is perhaps the most well-validated MRI measure in AD studies. This rank order refers to a situation in which all measures are used jointly for classification. Also, the gray highlighted measures are the ones that, when used jointly, gave the best classification accuracy in our independent test datasets (see **Figure 3.1** for ROC curves).

	Biomarker	
Rank	AD vs. control (weight / w_i^2)	MCI vs. control (weight / w_i^2)
1	MRI Hip[a] 0.1664	MRI Hip 0.1045
2	ApoE 0.1063	ApoE 0.0938
3	Age 0.0369	Age 0.0188
4	MRI Vent[b] 0.0349	MRI Vent 0.0103
5	MRI Temp[c] 0.0210	MRI Temp 0.0045
6	BMI 0.0147	BMI 0.0019
7	Sex 0.0013	Sex 0.0009

[a]Hippocampal volume summary
[b]Ventricular volume summary
[c]Temporal lobe summary from tensor-based morphometry (TBM)
Groups of biomarkers yielding the highest leave-one-out accuracy are highlighted.

We then aimed to find the top *N* (*N* ranging from 1 to 7) features that yielded the highest leave-one-out accuracy in the training set, using an RBF kernel with parameter optimization. Both linear and RBF kernels identified the same set of top features, but the RBF kernel gave better performance, so we only present those results here. For AD vs. control, the best combination included the top 4 features (baseline hippocampal and ventricular volumes, as well as ApoE and age); this joint classifier yielded a leave-one-out accuracy of 82.21% correct classification, with a corresponding area under the ROC curve (AUC) of 0.945, which is relatively high. For classifying MCI vs. control, the best feature combination consisted of the top 3 (baseline hippocampal volume, ApoE and age), which gave 70.89% accurate classification, with a corresponding area under the ROC curve of 0.860. As expected, MCI classification accuracy was slightly poorer than AD classification, as there is substantial overlap on all known measures, between MCI and normal aging. The best biomarker sets for each classification are highlighted in **Table 3.2**. **Figure 3.1** shows the ROC curves**.** In **Table 3.2,** only a subset of features was actually used: the best classifiers did not include BMI, sex, and the TBM-derived numeric summary. Also in **Table 3.2,** it is interesting that ventricular volume was helpful for the AD classification problem but not for distinguishing MCI from controls. This is reasonable given past findings by ourselves and others that ventricular expansion in MCI is relatively mild; there is also substantial cross-subject variation in ventricular volume, even in healthy subjects (Chou, et al., 2009b), and this may throw off a classifier's accuracy unless the disease effect outweighs this natural variation (Chou, et al., 2008; Chou, et al., 2009a; Chou, et al., 2009b).

Figure 3.1. ROC curves for AD and MCI classification. These curves show the trade-off between specificity and sensitivity for classifiers that best distinguished MCI from controls (*red curve*) and AD from controls (*blue curve*). The AD classifier used 4 measures and the MCI classifier only used 3. These evaluations are based on finding the top set of features that yielded the highest leave-one-out accuracies on the training set. The curves gradually rise, meaning that there is a natural trade-off: the parameters of the classifier's decision boundary can be adjusted to be stricter or more lenient. For stricter classification settings, false positive classifications will decrease but so will the rate of true positives. Curves are slightly jagged and not perfectly smooth as they are based on a finite set of test data; with more data, they would be smoother.



B. Adding PET-FDG and CSF for multi-modality classification

In this study, our goal was to compare the predictive power from the best combination of features obtained above, which included MRI, ApoE, and age, to that obtained when also including the PET-FDG temporal summary and CSF biomarkers. This may seem like an artificial distinction between two lists of biomarkers, but, from a practical point of view, the first classifier could be applied to a study that only used MRI, while the extended classifier would also need PET scans and lumbar puncture to be performed. Although using more features is almost

certainly better statistically, we wanted to assess how much difference it made, given the added expense, logistics, and possible attrition effects of performing multiple assessments.

Here, we considered three subsets of the ADNI subjects (N=328 when adding CSF alone, N=364 when adding PET-FDG alone, and N=166 when adding both CSF and PET-FDG) of the ADNI subjects, for whom the data from these additional diagnostic modalities were available. We applied the same ranking algorithm based on SVM weights and obtained rank orders for the biomarkers, with CSF and PET-FDG taken into account. We found the top set of biomarkers yielding the highest leave-one-out accuracies on the training set for each classification. The rank orders and best sets of biomarkers are displayed in **Table 3.3**. CSF t-tau and $a\beta_{42}$ were included in the best set of biomarkers for both AD and MCI classification. PET-FDG also contributed substantially to AD and MCI classification. The remaining top biomarkers were essentially the same as the ones identified in the above study.

It may seem paradoxical that when we list the biomarkers in order of priority (**Table 3.3**) some of them are listed even though they are not ultimately used in the best-performing classifier (only the lists of features in gray are used in the best classifier). The reason this occurs is that when all features are included, some features are given non-zero weights, which means that they are useful for classifying the training set. Even so, these features may give no detectable improvement in classifying the test set, so they were dropped from the final classifier. This does not mean that they are not useful predictors under any circumstances; it just means that in this sample, they did not improve classification accuracy on the independent evaluation data.

Table 3.3. Rank order list with relative SVM weights for MRI, ApoE, Age, Sex, BMI and either (a) CSF or (b) PET-FDG, for AD and MCI classification. Biomarkers are ranked according to their relative weights (contributions) in an SVM classifier that includes them all. A secondary question is which subset of these gives best classification accuracy, and this sublist is shown in gray. In these sublists, some features are omitted as adding them does not improve classification accuracy. Of the CSF markers, p-tau is relatively unhelpful but both t-tau and $a\beta_{42}$ provide independent predictive value. PET-FDG is a useful feature; whether it ranks above MRI hippocampal measures or not depends on whether the task is MCI or AD classification (hippocampal volume is slightly more useful than PET for MCI). PET measures are also somewhat correlated with MRI measures, so that when they are both included, each absorbs some of the variance; this may explain why ApoE genotype rises to the top of the predictors in terms of its independent contribution when MRI and PET are both included (*last two columns*).

	Biomarker			
	a. MRI + CSF		b. MRI + PET-FDG	
Rank	AD vs. control (weight / w_i^2)	MCI vs. control (weight / w_i^2)	AD vs. control (weight / w_i^2)	MCI vs. control (weight / w_i^2)
1	MRI Hip[a] 0.0794	MRI Hip 0.0519	ApoE 0.1529	ApoE 0.0929
2	CSF t-tau 0.0614	CSF $a\beta_{42}$ 0.0313	PET-FDG 0.1022	MRI Hip 0.0354
3	CSF $a\beta_{42}$ 0.0505	Age 0.0308	MRI Hip 0.0846	PET-FDG 0.0289
4	ApoE 0.0268	ApoE 0.0292	MRI Vent 0.0181	Age 0.0161
5	MRI Vent[b] 0.0238	CSF t-tau 0.0231	Age 0.0080	MRI Temp 0.0075
6	Age 0.0210	Sex 0.0157	MRI Temp 0.0057	Sex 0.0036
7	MRI Temp[c] 0.0163	MRI Temp 0.0085	BMI 0.0010	MRI Vent 0.0021
8	BMI 0.0077	BMI 0.0017	Sex 0.0004	BMI 0.0020
9	CSF p-tau 0.0003	CSF p-tau 0.0014		
10	Sex 0.0001	MRI Vent 0.0013		

[a]Hippocampal volume summary
[b]Ventricular volume summary
[c]Temporal lobe summary from tensor-based morphometry (TBM)
Sets of biomarkers yielding the highest leave-one-out accuracy are highlighted.

We then compared the performance of AD and MCI classifiers trained with the top biomarkers from the previous (N=635) study to those trained with the top biomarkers that included either CSF or the PET-FDG temporal summary or with both combined. Comparison of leave-one-out accuracies on the training set improved classification, implying that PET-FDG and CSF provide complementary information to MRI, ApoE and age. Leave-one-out accuracies for AD vs. control improved by 6.4%, 3.8%, and 11.6% by adding CSF alone, PET alone and both CSF and PET respectively. The corresponding improvements for MCI vs. control were 2.3%, 2.7%, and 4.6%.

When we compared the ROC curve AUCs, however, the improvement obtained by adding CSF, PET-FDG or both measures to the MRI measures was not statistically significant (p values > 0.05; **Table 3.4**). This lack of statistical significance may be due to the small size of the testing sets. If, however, this lack of significance is verified in even larger studies, it could have considerable implications for clinical trials in terms of total cost, efficiency and adverse effects.

Table 3.4. Comparison of AD and MCI classification accuracy and false positive/false negative trade-offs (ROC analyses) for classifiers that use different types of information: MRI, MRI+CSF, MRI+PET-FDG, and MRI+CSF+PET-FDG. Information for the top MRI classifier is listed twice, because MRI data were available for all ADNI subjects, but CSF and PET data were available only for a subset of those who had MRI. So it is only fair to report the classification accuracy on the full sample of MRIs, as well as on the subsamples in which head-to-head comparisons could be made with classifiers that also included the available CSF and PET data. The classifiers include ApoE and age, but not sex or BMI as the latter two did not contribute to the classification accuracies.

	AD versus control			MCI versus control		
Biomarkers	**LOOCV Accuracy**	**ROC AUC ± SE**	**Δ AUC[a] (p value)**	**LOOCV Accuracy**	**ROC AUC ± SE**	**Δ AUC (p value)**
Top MRI[b]	0.8160	0.8940 ± 0.0499	-	0.8421	0.8350 ± 0.0632	-
Top MRI+CSF[c]	0.8800	0.9560 ± 0.0273	0.0620 (0.191)	0.8647	0.8125 ± 0.0672	0.0225 (0.722)
Top MRI[b]	0.7634	0.7760 ± 0.0585	-	0.7227	0.7067 ± 0.0696	-
Top MRI+PET-FDG[d]	0.8011	0.7820 ± 0.0580	0.0060 (0.906)	0.7500	0.7444 ± 0.0672	0.0377 (0.382)
Top MRI[b]	0.7907	0.8850 ± 0.0501	-	0.7121	0.7488 ± 0.0649	-
Top MRI+CSF+PET-FDG[e]	0.9070	0.9175 ± 0.0413	0.0325 (0.357)	0.7576	0.7688 ± 0.0669	0.0200 (0.709)

[a]AUC difference relative to using the top MRI-based classifier only
[b]Top biomarkers identified in the N=635 study with MRI
[c]Top biomarkers identified in the N=328 study with MRI and CSF
[d]Top biomarkers identified in the N=364 study with MRI and PET-FDG
[e]Top biomarkers identified in the N=166 study with MRI, CSF and PET-FDG
LOOCV: leave-one-out cross-validation

C. Boosting Power for Clinical Trials

A novel use of classifiers is to identify subjects who are more likely to decline. Under some reasonable assumptions (see Discussion), this can lead to larger effect sizes for detecting changes in biomarkers over time; this may also be useful for reducing sample size requirements for clinical trials of potential disease-modifying therapies. In the past, several authors have suggested that people in the lowest 50% (or some other quantile) of hippocampal volume are more likely to show future decline, both clinically (e.g., conversion from MCI to AD) and on imaging (see, e.g., Frisoni et al, 2010). Of course, this idea could be generalized to defining a sample based on the *k*% of subjects that a classifier declares as most likely to decline clinically in

the future. Such a classifier could include not just MRI but any biomarker relevant for improving prediction.

As such, we computed minimum sample size estimates (n80) for the top *k* percent of subjects (for different values of *k* noted below) classified as *most likely to have AD* with our best AD classifier, using MRI hippocampal and ventricular summaries, ApoE and age as features. This *k*% of people are subjects in the independent test datasets (not used to train the classifier) who are assigned by the classifier to the AD class; they are those classified as AD who are farthest from the "SVM classifier decision boundary". We did not include PET-FDG and CSF biomarkers here, since adding these covariates limited our sample size and, as shown above, did not significantly improve classification in our tests. The subjects were ranked based on the SVM classifier output, the arithmetic sign of which determines the class assigned to each subject. A few AD subjects were excluded from the training and testing sets to avoid any overlap with the training set used in our prior report (Hua, et al., 2009) for creating the statistical ROIs. The results are shown in **Figure 3.2a.** When *k* is less than about 33%, the power estimates for AD subjects are improved compared to the minimal sample size of 48 AD subjects reported by Hua et al. (2009). There is a drop in the sample sizes needed to show a specific slowing effect, as the more AD-like subjects are selected. This has to be weighed against other factors (see Discussion), but it is interesting that the changes in these subjects have a greater effect size. It is also by no means obvious in advance that these subjects would give greater effect sizes. For the effect size to be greater, the changes have to be large and their variance has to be small; restricting the sample did not lead to an increase in the variability of the change measures sufficient to deplete effect sizes.

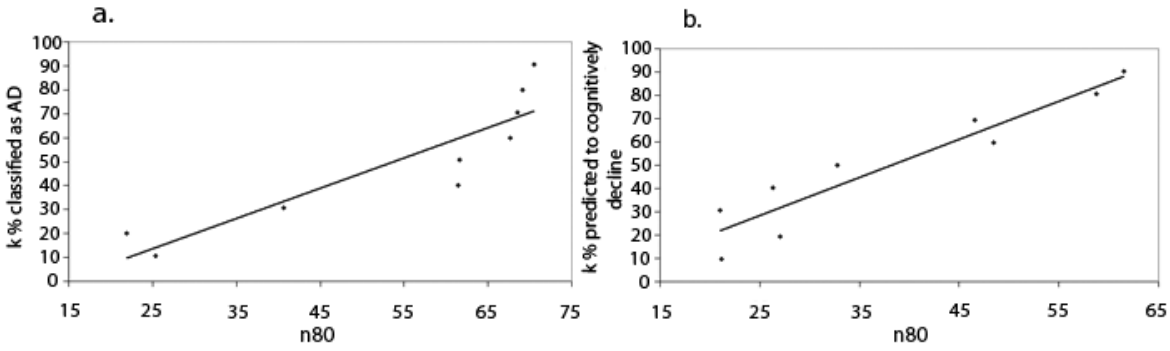
We could use the classifiers in many different ways to define a subsample – the *diagnostic* classifiers single out those who are most likely, based on all their imaging measures, to fall into a specific diagnostic category (e.g., AD). We also tested the benefit of defining a subsample of subjects with a classifier trained to identify likely decliners, based on all their imaging measures and other biomarkers, all at baseline.

To obtain similar n80 estimates for MCI subjects predicted to undergo cognitive decline, we considered 64 MCI subjects using MRI measures (3 features), PET-FDG, CSF biomarkers (3 features), ApoE and age. Here, the output of the SVM algorithm was set to be the 12-month rate of change in sobCDR, instead of a binary output for the classification approach used in the studies above. Training with all possible 2^9-1 feature combinations using a linear kernel (parameter C = 1) revealed PET-FDG, MRI ventricular and temporal summaries, and ApoE as the best set of features, with the lowest mean squared error on the testing set.

To increase our sample size for evaluating this classifier, we considered a larger group of 129 MCI subjects with only the four features identified above and trained a model that predicted the rate of sobCDR change in a novel testing set. We ranked the testing MCI subjects in order of predicted cognitive decline and computed n80 estimates for the top k% percent (for different values of k) of MCI subjects who the classifier predicted to be most likely to decline within a year (**Figure 3.2b**). The n80 values were even lower than the 88 MCI individuals we reported before as the minimal sample size for MCI (Hua, et al., 2009). In addition to sample size estimates reported by Hua et al. (2009), similar estimates have also been made by other

investigators such as Fox, et al. (2000), Jack, et al. (2004) and Schuff, et al. (2009), and we were able to improve upon these too with our approach.

Figure 3.2. n80 estimates (i.e., sample sizes required to detect a 25% slowing of the rate of atrophy with 80% power) as a function of restricting the sample to likely decliners. (a) Samples are based on the top *k*% classified, based on all biomarkers, as most likely to have AD (lower *k* gives smaller samples). (b) Here samples are based on the top *k*% of MCI subjects predicted by the classifier as most likely to decline (again lower values of *k* give dramatically lower samples). If only one-third of the most likely decliners were kept, in a sub-analysis based on the classifier's predictions, then the sample size needed (n80) for an MCI trial would only be around 30 subjects per arm (see Discussion for caveats of this approach).



In this report, we have considered AD and MCI classification as well as prediction of MCI conversion. Classifiers can also be trained to distinguish MCI converters from diagnostic groups other than MCI. For instance, when we performed classification with a small group of 12 MCI converters versus 12 healthy controls using all features in our study, we obtained a reasonably promising 71% accuracy, as this discrimination is more challenging than separating AD patients from controls.

In general, however, we do not want to discriminate MCI decliners from groups other than MCI for the prediction of later decline. We assumed here that MCI diagnosis was given, and we aimed to predict who would decline within that group. Predicting decline in a mixed group of controls and MCIs is a little easier, as the knowledge that a person is MCI is already fairly good evidence

that future decline is likely. Because of that, we wanted to assess the specific additive value of neuroimaging markers once a person is diagnosed as MCI (and it is reasonably helpful).

Discussion

We explored the power of several baseline biomarkers for AD and MCI, used jointly for diagnostic classification and for predicting future (1-year) cognitive decline in MCI. We also showed how to apply the multi-modality classifiers to choose sub-samples of subjects for boosting power in clinical trials. We determined combinations of regional MRI numerical summaries with demographic variables and ApoE that best classified AD vs. control and MCI vs. control. The top set of complementary biomarkers for AD classification (when used together) were the MRI hippocampal volume summary (measured with the method of Morra et al., 2008), ApoE genotype, age and the MRI ventricular summary (measured with the method of Chou et al., 2009) in that order, resulting in an 82.21% accuracy, and an ROC AUC of 0.945, which is quite strong. Biologically, hippocampal atrophy and ventricular enlargement are established manifestations of AD pathology, and the two structures are routinely monitored via MRI for AD clinical trials (Frisoni, et al., 2010). ApoE and advancing age are also well-known risk factors for AD (Carlsson, et al., 2009), and age is associated with atrophic rates in ADNI (Hua, et al., 2010a). The best set of features identified agrees with the AD literature. The one exception is the MRI temporal lobe summary, which did not improve classification power. This is not entirely surprising as it is quite highly correlated with the other two measures of atrophy (hippocampal and ventricular volume), so it may not add very much independent information for diagnostic classification. As expected, MCI classification was less accurate, and ventricular summaries were not as helpful; the best MCI diagnostic classifier only used hippocampal volume, ApoE

genotype and age (Frisoni, et al., 2010; Petersen, 2010). When compared to accuracy results reported by groups such as Vemuri et al. (2008), Klöppel et al. (2008) and Fan et al. (2008), our accuracies may seem a bit low. Perhaps, the main reason the accuracy values are not so high is that we are using numerical summary measures (single values for each imaging modality) as opposed to voxel-wise maps (which are implemented in papers that report higher accuracies). Even so, it is difficult to compare the results across papers as different subject samples are used. For example, ADNI considers only AD patients with relatively mild AD, and classification of AD is clearly easier in cohorts with a greater proportion of more severely affected patients. Even so, a fair comparison of accuracies between our study and others is difficult, not only because of sample variation, but also due to the fact that we utilized numerical summaries for imaging modalities, instead of voxel-wise maps. Even so, if some future classifier performs better, it could also be used to boost power using the same subpopulation selection method shown here.

By separately adding CSF biomarkers and PET-FDG as covariates for classification, where available, we obtained new rank order lists. These demonstrated how much the additional diagnostic measures contributed to AD and MCI classification, at least with this type of classifier. Different classes of AD biomarkers have dynamic trajectories that are thought to be temporally ordered with respect to the progression of the disease; in general, markers of amyloid deposition are thought to rise earlier than markers of neurodegeneration detectable on MRI, and these in turn become abnormal before tests of clinical function (Braskie, et al., 2008; Jack, et al., 2010; Petersen, 2010; Protas, et al., 2010).

It is therefore plausible to expect classifiers to perform best with biomarkers that are maximally dynamic during the stages of disease being considered; measurement reproducibility and precision are important. The top feature lists are generally consistent with this hypothesis, as MRI contributes more strongly to AD classification, whereas PET-FDG and CSF biomarkers, particularly $a\beta_{42}$, play more important roles in MCI classification. The observation that CSF tau levels were more important for AD classification, and CSF $a\beta_{42}$ more contributory to MCI classification is also consistent with Jack et al.'s model, in which the dynamic range of $a\beta_{42}$ precedes that of tau in the progression of AD. ApoE is consistently included among the best biomarkers for both AD and MCI classification, which agrees with another component of the Jack et al. (2010) hypothesis, stating that carrying E4 alleles may shift the sequence of biomarker activities to earlier time points relative to the onset of overtly detectable clinical symptoms.

Predicting future decline in MCI subjects is more challenging than AD and MCI classification, as differences among MCI subjects are subtle. Instead of approaching this problem with a binary classifier, we adapted the algorithm to predict a continuous cognitive outcome, which is the 12-month change in sobCDR. The baseline PET-FDG temporal summary, MRI temporal and ventricular summaries, and ApoE, were the best predictors of future cognitive decline in MCI (assessed over a 1-year follow-up interval). The combination of PET-FDG and ApoE genotype has been previously shown to provide good accuracy for predicting MCI conversion (Mosconi, et al., 2004). MRI-based temporal and ventricular volumes have also been reported for their predictive power in MCI subjects (Fleisher, et al., 2008; Korf, et al., 2004). It is mechanistically reasonable for this combination of structural, functional and genetic information to supply complementary predictive power. By using a multi-modality regression approach to predicting

cognitive decline in ADNI subjects, a very recent study found that a linear combination of MRI and PET-FDG was a better predictor of cognitive decline than CSF biomarkers (Walhovd, et al., 2010), consistent with our best set of biomarkers. Unexpectedly, however, the MRI hippocampal summaries were not incorporated into our predictive model, which is surprising as hippocampal volume can be useful for prediction of MCI progression to AD (Apostolova, et al., 2006a; Apostolova, et al., 2006b; Apostolova, et al., 2007; Frisoni, et al., 2010). The presence of detectable extra-hippocampal atrophy (e.g. in the ventricles and white matter) may also be good predictors of whether an MCI patient is deteriorating.

Our choice of brain regions and imaging measures to analyze was based on discussions among the ADNI Clinical, MRI and PET Cores. We chose imaging measures that had been used successfully in the past for disease classification or to monitor disease progression, preferring those measures that could be derived efficiently from a large dataset, without substantial manual interaction with the images. Clinical ratings were based on those widely used in clinical trials (CDR and MMSE), and the CSF biomarker measures were those found to be most promising in pilot studies (Shaw, et al., 2009). Needless to say, more brain regions or alternative cognitive tests could be proposed, and could be added to those analyzed here to boost performance even further. Specifically, in conference abstracts, Alexander, et al. (2008) and Zhang, et al. (2008) have advocated a *multivariate network analysis* in which a very large number of regional brain volumes are jointly used as predictors, in an SVM model. Other groups have parcellated the brain into a large number of subregions, but found that temporal lobe regions showed the greatest disease-related changes and significantly outperformed any of the clinical or cognitive measures examined for both AD and MCI (Holland, et al., 2009). To single out brain regions that are most

promising for analysis of disease-related brain change, we also focused on pre-selecting voxels in maps of brain change that show greatest effect sizes in independent samples. We and others have found that a classifier can be given and entire brain image, and from it can derive the voxels whose signals are most promising for group classification (Sun, et al., 2009). By comparing different imaging measures (voxel-based, ROI-based, or surface-based; Gutman, et al., 2008), and different classifiers (SVM versus others), future studies may be able to gauge which aspects of the classifier (its mathematical design or the features used) are most relevant for boosting performance.

In addition to scanning all the subjects with MRI at 1.5 T field strength, one quarter of ADNI's subjects also received 3 T scans. In prior work (Ho et al., 2009), we studied 110 ADNI subjects scanned longitudinally at both 3 and 1.5 T, across a one-year interval. Our power analyses found that 37 AD and 108 MCI subjects would be needed at 1.5 T versus 49 AD and 166 MCI subjects at 3 T, to detect a 25% slowing of atrophy with 80% power, but these estimates did not differ significantly with field strengths. At both field strengths, temporal lobe atrophy rates were highly correlated with interval decline in Alzheimer's Disease Assessment Scale-cognitive subscale (ADAS-cog), mini-mental status exam (MMSE), and Clinical Dementia Rating sum-of-boxes (sobCDR) scores. To avoid modeling the effects of scanner field strength as a confound, here we used the 1.5T ADNI data only. Some additional work may be needed to show that 3T scanners perform equally well for all biomarkers assessed here. The few ADNI studies that have assessed the field strength effect (Ho et al., 2009; Kruggel et al., 2010) suggest that 1.5 and 3 T scanners did not significantly differ in their power to detect neurodegenerative changes over a year.

Some clinical measures, such as the Clinical Dementia Rating (sobCDR), were not used as features for classification to avoid circular inference. Since these measures are used in making a diagnosis, it would be circular to incorporate them into our diagnostic classifiers and then test their empirical accuracy relative to the diagnosis given by physicians in the clinic. Even so, if used in practice to assist diagnosis, a classifier could use more cognitive measures – including those conventionally used for diagnosis and any other relevant information. Even so, the diagnostic accuracy of such a classifier could not then be "independently" validated in the same way as we did here. Doing so would require some other form of independent diagnostic ground truth, not used by the classifier, such as autopsy confirmation of characteristic signs of AD neuropathology. This could in principle be done, but neuropathology is not available in large numbers for the ADNI cohort.

A major clinical application of disease classifiers is for boosting power for clinical trial by reducing sample size estimates required to observe therapeutic effects. The idea of targeting a subgroup for analysis of treatment effects is not new (Frisoni, et al., 2010). In fact, a drug trial for prodromal AD is currently recruiting subjects, with an inclusion criterion based on CSF $a\beta_{42}$ and t-tau (http://clinicaltrials.gov/ct2/show/NCT00890890?term=bms+alzheimer%27s&rank=2). It appears new, however, to base the selection on a machine learning-based classifier that combines numerous biomarkers, which include neuroimaging measures. Combinations of disease markers are more likely to achieve sample size reductions than using single measures, such as subpopulation selection based on hippocampal volume only (of course statistical power must be traded off against the logistical complexity and cost of collecting and analyzing multiple biomarker assessments). When we considered the subset of subjects classified as most likely to

have AD by our multi-feature AD classifier, and the most likely decliners in MCI, we were able to reduce the n80 estimates to fewer than 40 subjects for both AD and MCI, improving on those estimates we reported before (Ho, et al., 2009; Hua, et al., 2009; Hua, et al., 2010b). This result supports the concept of clinical trial enrichment, which has been previously advocated (Cummings, et al., 2007; Frisoni, et al., 2010; Hampel and Broich, 2009). Our enrichment strategy works because the subpopulation of subjects who are more likely to decline are selected based on disease classifiers and outcome predictors that integrate information from a number of complementary biomarkers.

We chose to compute sample sizes needed to detect a 25% slowing of atrophy with 80% power. While 25% is a reasonable target for a treatment that aims to slow atrophy, the exact number chosen is arbitrary. It is simple to compute sample size estimates for other percentage reductions in the atrophic rate, such as 5% or 50%, for example. As we noted in Hua, et al. (2010b), treatments may slow atrophy to different degrees, which may be denoted by k%, for different k. The sample size estimates required to detect a k% slowing of atrophy may be easily derived by multiplying the sample size estimates (n80) in this paper by $(25/k)^2$, as the numbers follow an inverse-square law. For example, 4 times as many subjects would be needed to detect a 12.5% slowing of atrophy (half of 25%), versus a 25% slowing of atrophy (Ho, et al., 2009). The quadratic relationship between the sample size estimates and the percentage atrophic rate is illustrated in (Hua, et al., 2010b). Similarly, the results of this paper can be easily translated to studies aiming to detect a different level of treatment effect, and our findings remain unaffected as multiplying the variables by a constant $(25/k)^2$ does not alter the ranking of the effect sizes in the statistical tests (it is a monotone transformation, i.e., it preserves the rank order).

As a caveat, the n80 "minimal sample size" measure is practical but has limitations: first, it is based on changes in the patient groups only, and not their difference from controls; second, it assumes that a treatment would slow atrophy in the same places as it normally occurs, with the clinical outcome as observing an untreated sample with less atrophy. Finally, any treatment effects in a sub-analysis might only apply to people who fit the selection criteria for that sub-analysis; even so, evidence of an effect in a sub-analysis might suffice to initiate a broader study.

The approach and results reported here are relevant to future work in the neuroimaging of AD in several ways. First, several authors advocate "enrichment" in clinical trials by trying to select those most likely to decline, based on clinical criteria, or occasionally based on imaging criteria. This can be done by applying thresholds or cut-offs to volumetric measures on MRI scans, such as hippocampal volume, but here we advocate using the full armory of imaging and CSF measures to classify subjects first, and then use the classifier's output to select subpopulations for later statistical testing.

Although this may seem like basing the statistical approach in part on the data collected, rather than specifying it all in advance of the study, this approach would identify subjects whose imaging data made them most likely to show treatment effects, regardless of the treatment. A similar approach to boost the power of imaging biomarkers is voxel-set pre-selection, which substantially boosts power to detect the slowing of atrophy (Hua, et al., 2010; Chen, et al., 2010).

For these statistically-guided measures to be widely adopted as outcome measures in clinical trials, there needs to be some flexibility on the part of regulatory bodies that some features of the data collected may play a role in establishing which measures or subjects are evaluated. The analysis strategy can then adapt to the incoming data, and can exploit the power of Bayesian statistics and machine learning to obtain more powerful measures. It is quite defensible - and even advisable - for these machine learning approaches to be used, so long as the independence of statistical training and test samples is rigorously maintained.

A limitation of our study is that sample sizes become small when multiple imaging modalities and biomarkers are considered. In longitudinal studies especially, assessments of many kinds bring added costs, complexity, logistical difficulty, subject burden, and subject attrition (although in ADNI, attrition rates are only around 7% per year). Larger cohorts of subjects with available data from multiple biomarkers would allow more powerful classifiers and predictors to be developed, incorporating the best combinations of available diagnostic tools. More accurate ranking of biomarkers for verifying the details of Jack et al.'s temporal sequence hypothesis would become feasible. In addition, future studies will include additional diagnostic modalities such as Pittsburgh compound B (PiB), diffusion tensor imaging (DTI), arterial spin labeling (ASL) and resting state functional MRI for disease classification. PiB has been collected in a small subsample of ADNI subjects, but we did not evaluate it here as requiring all biomarkers would have further limited our sample sizes. Another future direction would be to employ machine learning algorithms other than SVM (e.g., boosting; Morra et al., 2009b), or classifiers based on features in voxel-based maps (Sun, et al., 2009), to improve classification and prediction accuracy. More powerful classifiers may then be implemented to improve upon our

clinical trial boosting results. Furthermore, machine learning can perhaps be used to discover genetic (Stein, et al., 2009; Stein, et al., 2010), epidemiological and physiological factors that influence the progression of AD.

References

Alexander GE, Hanson KD, Chen K, Reiman EM, Bernstein MA, Kornak J, Schuff NW, Fox NC, Thompson PM, Weiner MW, Jack CR Jr. Six month MRI gray matter declines in Alzheimer's dementia evaluated by voxel-based morphometry with multivariate network analysis: Preliminary findings from the ADNI study. The International Conference on Alzheimer's Disease. Chicago, IL, USA. 2008.

Alzheimer's Association. Alzheimer's disease facts and figures. Alzheimers Dement. 2009; 5(3):234-270.

Apostolova LG, Dinov ID, Dutton RA, Hayashi KM, Toga AW, Cummings JL, Thompson PM. 3D comparison of hippocampal atrophy in amnestic mild cognitive impairment and Alzheimer's disease. Brain. 2006a; 129(11):2867-2873.

Apostolova LG, Dutton RA, Dinov ID, Hayashi KM, Toga AW, Cummings JL, Thompson PM. Conversion of mild cognitive impairment to Alzheimer disease predicted by hippocampal atrophy maps. Arch Neurol. 2006b; 63(5):693-699.

Apostolova LG, Steiner CA, Akopyan GG, Dutton RA, Hayashi KM, Toga AW, Cummings JL, Thompson PM. Three-dimensional gray matter atrophy mapping in mild cognitive impairment and mild Alzheimer disease. Arch Neurol. 2007; 64(10):1489-1495.

Azad NA, Al Bugami M, Loy-English I. Gender differences in dementia risk factors. Gend Med. 2007; 4(2):121-129.

Braskie MN, Klunder AD, Hayashi KM, Protas H, Kepe V, Miller KJ, Huang SC, Barrio JR, Ercoli LM, Siddarth P, Satyamurthy N, Liu J, Toga AW, Bookheimer SY, Small GW, Thompson PM. Plaque and tangle imaging and cognition in normal aging and Alzheimer's disease. Neurobiol Aging. 2008; doi:10.1016/j.neurobiolaging.2008.09.012.

Buchman AS, Wilson RS, Bienias JL, Shah RC, Evans DA, Bennett DA. Change in body mass index and risk of incident Alzheimer disease. Neurology 65:892-897. 2005.

Burges CJC. A Tutorial on Support Vector Machines for Pattern Recognition. Data Min Knowl Disc. 1998; 2:121-167.

Carlsson CM, Gleason CE, Puglielli L, Asthana S. Chapter 65. Dementia Including Alzheimer's Disease. in: Halter JB, Ouslander JG, Tinetti ME, Studenski S, High KP, Asthana S (Eds.). Hazzard's Geriatric Medicine and Gerontology, 6e. The McGraw-Hill Companies. URL: http://www.accessmedicine.com/content.aspx?aID=5122625. 2009.

Chen K, Langbaum JB, Fleisher AS, Ayutyanont N, Reschke C, Lee W, Liu X, Bandy D, Alexander GE, Thompson PM, Foster NL, Harvey DJ, de Leon MJ, Koeppe RA, Jagust WJ, Weiner MW, Reiman EM. Twelve-Month Metabolic Declines in Probable Alzheimer's Disease and Amnestic Mild Cognitive Impairment Assessed Using an Empirically Pre-

Defined Statistical Region-of-Interest: Findings from the Alzheimer's Disease Neuroimaging Initiative. NeuroImage. 2010; doi:10.1016/j.neuroimage.2010.02.064.

Chou YY, Lepore N, Avedissian C, Madsen SK, Parikshak N, Hua X, Shaw LM, Trojanowski JQ, Weiner MW, Toga AW, Thompson PM. Mapping correlations between ventricular expansion and CSF amyloid and tau biomarkers in 240 subjects with Alzheimer's disease, mild cognitive impairment and elderly controls. NeuroImage. 2009a; 46(2):394-410.

Chou YY, Lepore N, Chiang MC, Avedissian C, Barysheva M, McMahon KL, de Zubicaray GI, Meredith M, Wright MJ, Toga AW, Thompson PM. Mapping genetic influences on ventricular structure in twins. NeuroImage. 2009b; 44(4):1312-1323.

Chou YY, Lepore N, de Zubicaray GI, Carmichael OT, Becker JT, Toga AW, Thompson PM. Automated ventricular mapping with multi-atlas fluid image alignment reveals genetic effects in Alzheimer's disease. NeuroImage. 2008; 40(2):615-630.

Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, Roses AD, Haines JL, Pericak-Vance MA. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science. 1993; 261(5123):921-923.

Cummings JL, Doody R, Clark C. Disease-modifying therapies for Alzheimer disease: challenges to early intervention. Neurology. 2007; 69(16):1622-1634.

Davatzikos C, Resnick SM, Wu X, Parmpi P, Clark CM. Individual patient diagnosis of AD and FTD via high-dimensional pattern classification of MRI. NeuroImage. 2008; 41(4):1220-1227.

Davatzikos C, Xu F, Yang A, Yong F, Resnick SM. Longitudinal progression of Alzheimer's-like patterns of atrophy in normal older adults: the SPARE-AD index. Brain. 2009; 132(8):2026-2035.

Duara R, Loewenstein DA, Potter E, Appel J, Greig MT, Urs R, Shen Q, Raj A, Small B, Barker W, Schofield E, Wu Y, Potter H. Medial temporal lobe atrophy on MRI scans and the diagnosis of Alzheimer disease. Neurology. 2008; 71, 1986-1992.

Ecker C, Rocha-Rego V, Johnston P, Mourao-Miranda J, Marquand A, Daly EM, Brammer MJ, Murphy C, Murphy DG. Investigating the predictive value of whole-brain structural MR scans in autism: A pattern classification approach. NeuroImage. 2010; 49(1):44-56.

Fan Y, Resnick SM, Wu X, Davatzikos C. Structural and functional biomarkers of prodromal Alzheimer's disease: a high-dimensional pattern classification study. NeuroImage. 2008; 41(2):277-285.

Fleisher AS, Sun S, Taylor C, Ward CP, Gamst AC, Petersen RC, Jack CR Jr, Aisen PS, Thal LJ. Volumetric MRI vs clinical predictors of Alzheimer disease in mild cognitive impairment. Neurology. 2008; 70(3):191-199.

Fox NC, Cousens S, Scahill R, Harvey RJ, Rossor MN. Using serial registered brain magnetic resonance imaging to measure disease progression in Alzheimer disease. Arch Neurol. 2000; 57:339-344.

Freund Y, Schapire RE. A Short Introduction to Boosting. J Japan Soc for Artif Intell. 1999; 14(5):771-780.

Frisoni GB, Fox NC, Jack CR Jr, Scheltens P, Thompson PM. The clinical use of structural MRI in Alzheimer disease. Nat Rev Neurol. 2010; 6:1-11.

Gutman B, Wang YL, Morra JH, Tu Z, Jack CR Jr, Weiner MW, Toga AW, Thompson PM. Disease Classification with Hippocampal Surface Invariants. MICCAI Workshop on Hippocampal Mapping. 2008.

Hampel H, Broich K. Enrichment of MCI and early Alzheimer's disease treatment trials using neurochemical & imaging candidate biomarkers. J Nutr Health Aging. 2009; 13(4):373-375.

Hanley JA, McNeil BJ. A Method of Comparing the Areas under Receiver Operating Characteristic Curves Derived from the Same Cases. Radiology. 1983; 148(3):839-43.

Hinrichs C, Singh V, Mukherjee L, Xu G, Chung MK, Johnson SC. Spatially augmented LPboosting for AD classification with evaluations on the ADNI dataset. NeuroImage. 2009; 48(1):138-149.

Ho AJ, Stein JL, Hua X, Lee S, Hibar DP, Leow AD, Dinov ID, Toga AW, Saykin AJ, Shen L, Foroud T, Pankratz N, Huentelman MJ, Craig DW, Gerber JD, Allen A, Corneveaux J, Stephan DA, Webster J, DeChairo BM, Potkin SG, Jack CR Jr, Weiner MW, Raji CA, Lopez OL, Becker JT, Thompson PM. Commonly carried allele within FTO, an obesity-associated gene, relates to accelerated brain degeneration in the elderly, Proc Natl Acad Sci USA. submitted. 2010a.

Ho AJ, Raji CA, Becker JT, Lopez OL, Kuller LH, Hua X, Lee S, Hibar D, Dinov ID, Stein JL, Jack CR Jr, Weiner MW, Toga AW, Thompson PM. Obesity and brain structure in 700 MCI and AD patients. Neurobiol Aging. submitted. 2010b.

Ho AJ, Hua X, Lee S, Yanovsky I, Leow AD, Gutman B, Dinov ID, Toga AW, Jack CR Jr, Bernstein MA, Reiman EM, Harvey D, Kornak J, Schuff N, Alexander GE, Weiner MW, Thompson PM. Comparing 3T and 1.5T MRI for tracking AD progression with tensor-based morphometry. Hum Brain Mapp. 2009; doi:10.1002/hbm.20882.

Holland D, Brewer JB, Hagler DJ, Fenema-Notestine C, Dale AM. Subregional neuroanatomical change as a biomarker for Alzheimer's disease. Proc Natl Acad Sci USA. 2009; 106(49):20954-20959.

Hua X, Hibar DP, Lee S, Toga AW, Jack CR Jr, Weiner MW, Thompson PM. Sex and age differences in atrophic rates: an ADNI study with N=1368 MRI scans. Neurobiol Aging. submitted. 2010a.

Hua X, Lee S, Hibar DP, Yanovsky I, Leow AD, Toga AW, Jack Jr Cr, Bernstein MA, Reiman EM, Harvey DJ, Kornak J, Schuff N, Alexander GE, Weiner MW, Thompson PM. Mapping Alzheimer's disease progression in 1309 MRI scans: power estimates for different inter-scan intervals. NeuroImage. submitted. 2010b.

Hua X, Lee S, Yanovsky I, Leow AD, Chou YY, Ho AJ, Gutman B, Toga AW, Jack CR Jr, Bernstein MA, Reiman EM, Harvey DJ, Kornak J, Schuff N, Alexander GE, Weiner MW, Thompson PM. Optimizing power to track brain degeneration in Alzheimer's disease and mild cognitive impairment with tensor-based morphometry: an ADNI study of 515 subjects. NeuroImage. 2009; 48(4):668-681.

Hua X, Leow AD, Lee S, Klunder AD, Toga AW, Lepore N, Chou YY, Brun C, Chiang MC, Barysheva M, Jack CR Jr, Bernstein MA, Britson PJ, Ward CP, Whitwell JL, Borowski B, Fleisher AS, Fox NC, Boyes RG, Barnes J, Harvey D, Kornak J, Schuff N, Boreta L, Alexander GE, Weiner MW, Thompson PM. 3D characterization of brain atrophy in Alzheimer's disease and mild cognitive impairment using tensor-based morphometry. NeuroImage. 2008a; 41(1):19-34.

Hua X, Leow AD, Parikshak N, Lee S, Chiang MC, Toga AW, Jack CR Jr, Weiner MW, Thompson PM. Tensor-based morphometry as a neuroimaging biomarker for Alzheimer's disease: an MRI study of 676 AD, MCI, and normal subjects. NeuroImage. 2008b; 43(3):458-69.

Jack CR Jr, Knopman DSK, Jagust WJ, Shaw LM, Aisen PS, Weiner MW, Petersen RC, Trojanowski JQ. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. Lancet Neurology. 2010; 9:119-128.

Jack CR Jr, Shiung MM, Gunter JL, O'Brien PC, Weigand SD, Knopman DS, Boeve BF, Ivnik RJ, Smith GE, Cha RH, Tangalos EG, Petersen RC. Comparison of different MRI brain atrophy rate measures with clinical disease progression in AD. Neurology. 2004; 62(4):591-600.

Klöppel S, Stonnington CM, Chu C, Draganski B, Scahill RI, Rohrer JD, Fox NC, Jack CR Jr, Ashburner J, Frackowiak RSJ. Automatic classification of MR scans in Alzheimer's disease. Brain. 2008; 131:681-689.

Klunk WE, Engler H, Nordberg A, Wang Y, Blomqvist G, Holt DP, Bergstrom M, Savitcheva I, Huang GF, Estrada S, Ausen B, Debnath ML, Barletta J, Price JC, Sandell J, Lopresti BJ, Wall A, Koivisto P, Antoni G, Mathis CA, Langstrom B. Imaging brain amyloid in Alzheimer's disease with Pittsburgh Compound-B. Ann Neurol. 2004; 55(3):306-319.

Korf ES.C, Wahlund LO, Visser PJ, Scheltens P. Medial temporal lobe atrophy on MRI predicts dementia in patients with mild cognitive impairment. Neurology. 2004. 63, 94-100.

Koutsouleris N, Meisenzahl EM, Davatzikos C, Bottlender R, Frodl T, Scheuerecker J, Schmitt G, Zetzsche T, Decker P, Reiser M, Möller HJ, Gaser C. Use of Neuroanatomical Pattern Classification to Identify Subjects in At-Risk Mental States of Psychosis and Predict Disease Transition. Arch Gen Psychiatry. 2009; 66(7):700-712.

Kruggel F, Turner J, Muftuler LT. Impact of scanner hardware and imaging protocol on image quality and compartment volume precision in the ADNI cohort. NeuroImage. 2010; 49(3):2123-2133.

Landau SM, Harvey D, Madison CM, Koeppe RA, Reiman EM, Foster NL, Weiner MW, Jagust WJ. Associations between cognitive, functional, and FDG-PET measures of decline in AD and MCI. Neurobiol Aging. 2009; doi:10.1016/j.neurobiolaging.2009.07.002.

Lerch JP, Pruessner J, Zijdenbos AP, Collins DL, Teipel SJ, Hampel H, Evans AC. Automated cortical thickness measurements from MRI can accurately separate Alzheimer's patients from normal elderly controls. Neurobiol Aging. 2008; 29(1):23-30.

Lindsay J, Laurin D, Verreault R, Hébert R, Helliwell B, Hill GB, McDowell I. Risk Factors for Alzheimer's Disease: A Prospective Analysis from the Canadian Study of Health and Aging. Am J Epidemiol. 2002; 156(5):445-453.

Lukas L, Devos A, Suykens JA.K, Vanhamme L, Howe FA, Majós C, Moreno-Torres A, Van Der Graaf M, Tate AR, Arús C, Van Huffel S. Brain tumor classification based on long echo proton MRS signals. Artif Intell Med. 2004; 31(1):73-89.

Mesrob L, Magnin B, Colliot O, Sarazin M, Hahn-Barma V, Dubois B, Gallinari P, Lehéricy S, Kinkingnéhun S, Benali H. Identification of Atrophy Patterns in Alzheimer's Disease Based on SVM Feature Selection and Anatomical Parcellation. Lecture Notes in Computer Science. 2008; 5128, 124-132.

Morra JH, Tu Z, Apostolova LG, Green AE, Avedissian C, Madsen SK, Parikshak N, Hua X, Toga AW, Jack CR Jr, Weiner MW, Thompson PM. Validation of a fully automated 3D hippocampal segmentation method using subjects with Alzheimer's disease mild cognitive impairment, and elderly controls. NeuroImage. 2008; 43(1):59-68.

Morra JH, Tu Z, Apostolova LG, Green AE, Avedissian C, Madsen SK, Parikshak N, Toga AW, Jack CR Jr, Schuff N, Weiner MW, Thompson PM. Automated mapping of hippocampal

atrophy in 1-year repeat MRI data from 490 subjects with Alzheimer's disease, mild cognitive impairment, and elderly controls. NeuroImage. 2009; 45(1 Suppl):S3-15.

Morra JH, Tu Z, Apostolova LG, Green AE, Toga AW, Thompson PM. Comparison of AdaBoost and support vector machines for detecting Alzheimer's disease through automated hippocampal segmentation. IEEE Trans Med Imaging. 2010; 29(1):30-43.

Mosconi L, Perani D, Sorbi S, Herholz K, Nacmias B, Holthoff V, Salmon E, Baron J-C, De Cristofaro MT.R, Padovani A, Borroni B, Franceschi M, Bracco L, Pupi A. MCI conversion to dementia and the APOE genotype, A prediction study with FDG-PET. Neurology. 2004; 63:2332-2340.

Mourão-Miranda J, Bokde AL.W, Born C, Hampel H, Stetter M. Classifying brain states and determining the discriminating activation patterns: Support Vector Machine on functional MRI data. NeuroImage. 2005; 28(4):980-995.

Petersen RC. Alzheimer's disease: progress in prediction. Lancet Neurol. 2010; 9(1):4-5.

Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: clinical characterization and outcome. Arch Neurol. 1999; 56(3):303-308.

Protas HD, Huang SC, Kepe V, Hayashi K, Klunder A, Braskie MN, Ercoli L, Bookheimer S, Thompson PM, Small GW, Barrio JR. FDDNP binding using MR derived cortical surface maps. NeuroImage. 1999; 49(1):240-248.

Raji CA, Ho AJ, Parikshak N, Becker JT, Lopez OL, Kuller LH, Hua X, Leow AD, Toga AW, Thompson PM. Brain Structure and Obesity. Hum Brain Mapp. 2008; doi:10.1002/hbm.20870.

Schuff N, Woerner N, Boreta L, Kornfield T, Shaw LM, Trojanowski JQ, Thompson PM, Jack CR Jr, Weiner MW. MRI of hippocampal volume loss in early Alzheimer's disease in relation to ApoE genotype and biomarkers. Brain. 2009; 132(4):1067-1077.

Shaw LM, Vanderstichele H, Knapik-Czajka M, Clark CM, Aisen PS, Petersen RC, Blennow K, Soares H, Simon A, Lewczuk P, Dean R, Siemers E, Potter W, Lee VM, Trojanowski JQ. Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. Ann Neurol. 2009; 65(4):403-413.

Stein JL, Hua X, Lee S, Ho AJ, Leow AD, Toga AW, Saykin AJ, Shen L, Foroud T, Pankratz N, Huentelman MJ, Craig DW, Gerber JD, Allen A, Corneveaux J, Stephan DA, Webster J, DeChairo BM, Potkin SG, Jack CR Jr, Weiner MW, Thompson PM. Voxelwise Genome-Wide Association Study (vGWAS). NeuroImage. submitted. 2009.

Stein JL, Hua X, Morra JH, Lee S, Hibar DP, Ho AJ, Leow AD, Toga AW, Sul JH, Kang H, Eskin E, Saykin AJ, Shen L, Foroud T, Pankratz N, Huentelman MJ, Craig DW, Gerber JD, Allen AN, Corneveaux JJ, Stephan DA, Webster J, DeChairo BM, Potkin SG, Jack CR Jr, Weiner MW, Thompson PM. Genome-wide association study of temporal lobe structure identifies novel quantitative trait loci for neurodegeneration in Alzheimer’s disease. NeuroImage. submitted. 2010.

Sun D, van Erp TG.M, Thompson PM, Bearden CE, Daley M, Kushan L, Hardt ME, Nuechterlein K, Toga AW, Cannon TD. Elucidating an MRI-based biomarker for psychosis: classification using probabilistic brain atlas and machine learning algorithms. Biol Psychiatry. 2009; doi:10.1016/j.biopsych.2009.07.019.

Suykens JA.K, Vandewalle J. Least Squares Support Vector Machine Classifiers. Neural Process Lett. 1999; 9:293-300.

Vapnik V. The Nature of Statistical Learning Theory. Springer, New York. 1995.

Vemuri P, Gunter JL, Senjem ML, Whitwell JL, Kantarci K, Knopman DS, Boeve BF, Petersen RC, Jack CR Jr. Alzheimer's disease diagnosis in individual subjects using structural MR images: validation studies. NeuroImage. 2008; 39(3):1186-1197.

Vemuri P, Wiste HJ, Weigand SD, Shaw LM, Trojanowski JQ, Weiner MW, Knopman DS, Petersen RC, Jack CR Jr. MRI and CSF biomarkers in normal, MCI, and AD subjects: predicting future clinical change. Neurology. 2009; 73(4):294-301.

Walhovd KB, Fjell AM, Brewer J, McEvoy LK, Fennema-Notestine C, Hagler DJ, JR, Jennings RG, Karow D, Dale AM. Combining MR Imaging, Positron-Emission Tomography, and CSF Biomarkers in the Diagnosis and Prognosis of Alzheimer Disease. AJNR Am J Neuroradiol. 2010; doi: 10.3174/ajnr.A1809.

Wilson SM, Ogar JM, Laluz V, Growdon M, Jang J, Glenn S, Miller BL, Weiner MW, Gorno-Tempini ML. Automated MRI-based classification of primary progressive aphasia variants. NeuroImage. 47(4):1558-1567. 2009.

Zhang H, Wu T, Bae M, Reiman EM, Alexander GE, Jack CR Jr, Thompson PM, Chen K. Use Of The Support Vector Machine And Sensitivity Of an AD-related Region-of-interest Gray Matter Classifier In Identifying Amnestic MCI Subjects Who Convert To AD: Preliminary Findings From The AD Neuroimaging Initiative. The International Conference on Alzheimer’s Disease. Chicago, IL, USA. 2008.

Chapter 4. Multilocus analysis of brain images

Introduction

Over the past decade, public and private funding institutions have invested billions of dollars in the fields of human neuroimaging and genetics (Akil et al., 2010). Recently, researchers have sought to use quantitative measures from brain images to test how genetic variation influences the brain. Imaging measures are thought to have a simpler genetic architecture than diagnostic measures based on cognitive or clinical assessments (Gottesman and Gould, 2003). In other words, the penetrance of an individual genetic polymorphism is expected to be higher at the imaging level than at the diagnostic level. As such, imaging-derived traits may offer more power to detect how specific genes contribute to brain disease. Genetic analysis of images has been used to discover how susceptibility genes affect brain integrity (Braskie et al., 2011b). Recent studies have revealed gene effects operating within an entire population, in the form of a 3D brain map (Thompson et al., 2001; Stein et al., 2010a; Hibar et al., 2011).

Optimally merging these two well-developed fields requires innovative mathematics and computational methods, guided by genomics and neuroscience. Imaging genetics is still a nascent field, and many studies are relatively simplistic - they generally test how a single genetic variant, or a small set of such variants (usually single nucleotide polymorphisms, or SNPs) are associated with a single summary measure of the brain. These studies begin to bridge the gap between the two fields, but do not take full advantage of advanced methods from either field, which can survey the entire genome or allow an image-wide search. By contrast, multivariate statistical methods such as machine learning and sparse regression can handle high dimensional

datasets. Many of these are being adapted to analyze a range of brain processes and biological markers of disease.

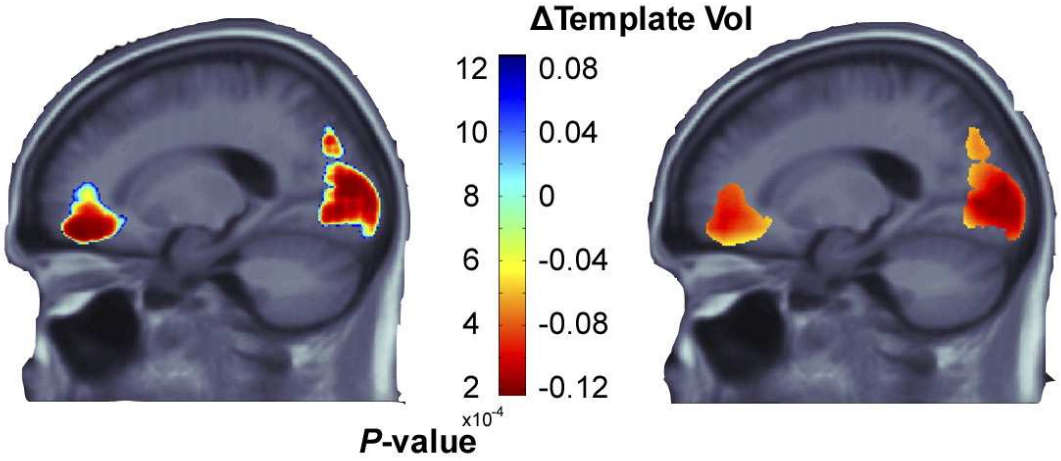
In this review, we summarize the recent evolution of imaging genetics, from candidate gene studies to multilocus methods and genome-wide searches to genome-wide, image-wide searches. We explain how images are used in different ways, ranging from single region-of-interest methods - that assess the volume or shape of a specific brain region, such as the hippocampus - to voxelwise approaches that survey the whole brain at once in 3D. In these efforts, multivariate, "multilocus" techniques can model how several genetic variants affect the brain at once. Specialized approaches - such as sparse coding methods - can simultaneously handle the high dimensionality and high degree of correlation observed across the genome and in image-derived maps.

Candidate Gene Studies

In studies that scan a large number of patients or controls, candidate gene studies have often been used to assess genetic effects on the brain. This approach is appealing as one can test biologically plausible hypotheses and determine how specific, well-studied genetic variations affect brain structure and function. Early studies, for instance, explored how genes related to serotonin transport affected measures extracted from single-photon emission computed tomography (SPECT) and functional magnetic resonance imaging (fMRI; Heinz et al., 2000; Hariri et al., 2002). Serotonin's role in neurotransmission and neuromodulation - and the well-known anatomy of the monoamine systems - made it possible to frame and confirm testable hypotheses for pertinent regions such as the raphe nuclei and amygdala (Munafo et al., 2008).

Candidate gene studies, such as those above, may assess a single measure derived from a specific region of interest (ROI) in the image. This may be the whole brain, or a subregion such as the gray matter, or the volume or mean activation of a subcortical region. More recently, voxel-by-voxel searches have been conducted to assess candidate gene effects throughout the whole brain in 3D. This unbiased search across the brain makes no prior assumptions on which regions may be affected. Statistical *maps* are also widely used in neuroimaging. Spatial statistics, such as PCA or ICA, may also be performed for dimension reduction, and multiple comparisons corrections, such as the false discovery rate method, can help to decide if a pattern of gene effects is significant across the voxels searched. For example, Ho et al. investigated the effects of a proxy SNP in the fat mass and obesity-associated (*FTO*) gene reliably associated with increased risk for obesity (rs3751812) (Frayling et al., 2007) on brain structure. They used MRI along with tensor-based morphometry, to evaluate 206 healthy elderly subjects. *FTO* risk allele carriers had lower frontal and occipital lobe volumes (**Figure 4.1**). In such studies, maps of statistical associations are created by performing separate association tests at each imaging voxel in the brain. As the number of statistical tests is very large, a standard correction for multiple comparisons can be used, such as the False Discovery Rate method (FDR; Benjamin and Hochberg, 1995) or its more advanced variants such as topological FDR (Chumbley et al., 2009), which consider the geometry of the effects. These corrections assess how likely it is that the overall pattern of associations could be observed by chance. Voxel-based analyses may also be informed by prior hypotheses: regions of interest may be defined as search regions, such as the temporal lobes, to include prior information on the expected location or patterns of effects (Stein et al., 2010a).

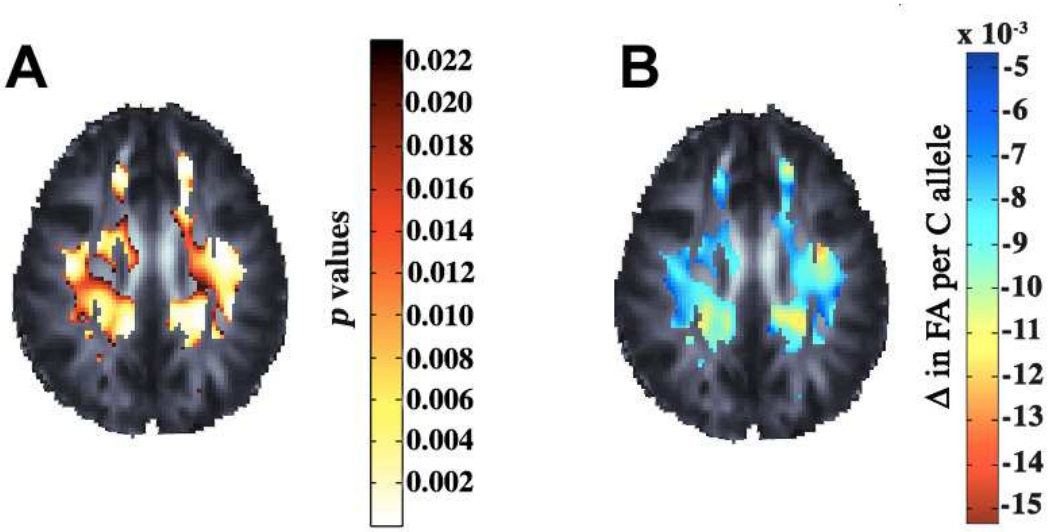
Figure 4.1. *P*-values (corrected using the false discovery rate method; *left panel*) and corresponding regression coefficients (*right*) show the statistical associations between a candidate single nucleotide polymorphism in the *FTO* gene (which is associated with higher risk of obesity) and tensor-based morphometry maps derived from anatomical MRI scans of the brain. Significant associations with regional brain volumes are detected in the occipital and frontal lobes. Clearly, if other regions had been specified in advance as the target of study, association effects may have been missed. Adapted from Ho et al. (2010).



Brain imaging measures used in genetic studies should ideally be highly heritable and be genetically related to a biological process affected by genetic variation, such as a disease process (Gottesman and Gould, 2003; Glahn et al., 2007; Winkler et al., 2010). Some argue that the use of imaging endophenotypes should boost power to detect genetic variants that have reliable but small effects on disease status (Meyer-Lindenberg and Weinberger, 2006). One neuroimaging modality that shows great promise in candidate gene studies is diffusion tensor imaging (DTI), which assesses the fiber integrity of the brain's white matter. DTI is based on the observation that myelination restricts water diffusion, and disease processes typically increase water diffusion across cell membranes (Beaulieu, 2002). Some DTI-derived measures, such as the fractional anisotropy of diffusion (FA), are widely accepted as measuring brain integrity. FA is highly heritable (Chiang et al., 2009; Kochunov et al., 2010) and is consistently altered in a range

of developmental and psychiatric disorders (Thomason and Thompson, 2011). Candidate polymorphisms already associated with brain disorders may be surveyed to discover associations with maps of DTI parameters such as fractional anisotropy. One recent DTI study of young healthy adults (Braskie et al., 2011a), studied the voxelwise effects of the rs11136000 SNP in the recently discovered Alzheimer's disease (AD) risk gene, *CLU*. Significant associations were detected in several anatomical regions that undergo atrophy in AD (**Figure 4.2**). In similar candidate gene studies using DTI, other genes such as *BDNF* (Chiang et al., 2011a) and *COMT* (Thomason et al., 2010) have been found to influence white matter structure, with carriers of one variant showing consistently higher or lower FA.

Figure 4.2. Corrected *p*-values (A) and regression coefficients (B) are shown for the voxelwise effects of a candidate polymorphism in the *CLU* gene – a highly prevalent Alzheimer's susceptibility gene – on fractional anisotropy maps derived from DTI scans of 398 young adults. The axial slice shows the extensive influence of the genetic variant on white matter structure. Adapted from Braskie et al. (2011a).



Genome-wide associations with single imaging measures

Candidate gene studies have successfully discovered patterns of brain differences associated with genetic variants whose function is relatively well known (such as ApoE, for example - a risk gene for late-onset Alzheimer's disease; Shaw et al., 2007). The choice of a candidate gene, however, requires a strong prior hypothesis, and most of the genetic determinants of the highly heritable imaging measures (connectivity or cortical thickness, for example) are unknown. In most candidate gene studies in imaging, there is a correction for multiple comparisons to control the rate of false discoveries across the image, but this does not take into account the genetic variant tested, or the fact that it could have been selected from a wide list of possibly associated genes. In genetics, and by extension imaging genetics, there is a high risk of false positive findings unless appropriate corrections are made. Moving beyond candidate gene studies to an unbiased search of the whole genome clearly requires an appropriate genome-wide significance criterion. Otherwise, many false positive associations will be reported that would not be replicated in the future (Ioannidis, 2005).

Genome-wide association (GWA) studies typically assess associations between hundreds of thousands of SNPs and a phenotype of interest (such as a disease, or a specific image-derived measure). GWA studies have discovered hundreds of common risk loci for diseases and traits in recent years (Hindorff et al., 2009). GWA studies are frequently conducted for discrete, case-control phenotypes, such as the diagnosis of a specific disease (such as Alzheimer's disease or schizophrenia vs. healthy control). These studies, however, are limited as participants do not always fall clearly into unique diagnostic categories, and may vary in dimensions not relevant to

disease (Pearson and Manolio, 2008). For neuropsychiatric disorders in particular, symptoms expressed by members of specific diagnostic groups may be highly heterogeneous – and there may also be substantial co-morbidity and overlap in symptom profiles across disorders (Psychiatric GWAS Consortium, 2009; Hall and Smoller, 2010).

Measures derived from brain images in principle are closer to the underlying biology of gene action, offering an alternative target for genome-wide searches, by serving as intermediate phenotypes or endophenotypes for GWA studies (Gottesman and Gould, 2003; Hall and Smoller, 2010). Several imaging GWA scans have been published: Potkin et al. (2009b) identified SNPs in two genes (*RSRC1* and *ARHGAP18*) that showed associations with a blood-oxygen-level dependent (BOLD) contrast measure from a brain region implicated in schizophrenia. Similarly, Stein et al. (2010a) discovered a SNP in the *GRIN2B* gene (rs10845840) and an intergenic SNP (rs2456930) associated with an MRI-derived tensor-based morphometry (TBM) measure of temporal lobe volume in 740 elderly subjects from the Alzheimer's Disease Neuroimaging Initiative. In these and other studies, linear regressions are used to assess the additive or dominant allelic effect of each SNP, after adjusting for covariates such as age and sex, and the confounding effects of population stratification (e.g., Potkin et al., 2009a). This yields *p*-values assessing the evidence for the association of each SNP with the imaging summary chosen. The overall significance of any one SNP effect is then assessed through a form of genome-wide correction for multiple comparisons. Commonly, a nominal *p*-value less than $5x10^{-8}$ is used.

The GWA study design has been extended to analyze whole images, but one of the shortcomings of all GWAS studies is their limited power (or alternatively, the large sample sizes needed) to

detect relevant gene variants. Most SNPs affecting the brain have modest effect sizes (often explaining <1% of the variance in a quantitative phenotype). Meta-analysis can provide added statistical power to discover variants with small effects. Replication, and meta-analysis in particular, have been widely embraced as a way to aggregate evidence from multiple genetic studies, including studies of disease risk, and normally-varying traits such as height (de Baaker et al., 2008; McCarthy et al., 2008; Zeggini and Ioannidis, 2009; Yang et al., 2010).

Even so, most imaging GWA studies consider under a thousand subjects, so are limited in detection power. This led many researchers in the field to band together to search for relevant genetic associations with imaging traits meta-analytically, in many large samples. One promising initiative is called Enhancing Neuro Imaging Genetics through Meta-Analysis (ENIGMA) and is currently accepting research groups who want to become involved in meta-analytic imaging genomics projects (http://enigma.loni.ucla.edu/). The ENIGMA pilot project is a large meta-analysis to discover genes associated with hippocampal volume on brain MRI in over 9,000 subjects scanned by 21 research centers (Stein et al., submitted). Future imaging genetics studies may rely on large meta-analyses and international collaborations to overcome the low power and relatively small effect sizes. However, some genetic associations can be found and replicated without vast meta-analytic approaches like ENIGMA. For example, Stein et al. (2011) discovered and replicated an association between caudate volume and the SNP rs163030 located in and around two genes, *WDR41* and *PDE8B*. These genes are involved in dopamine signaling and development; a Mendelian mutation in one leads to severe caudate atrophy. Similarly, Joyner et al., (2009) replicated an association with cortical surface area in a common variant

(rs2239464) of the *MECP2* gene, which is linked to microencephaly and other morphological brain disorders.

Genetic Analysis of Mass Univariate Imaging Phenotypes

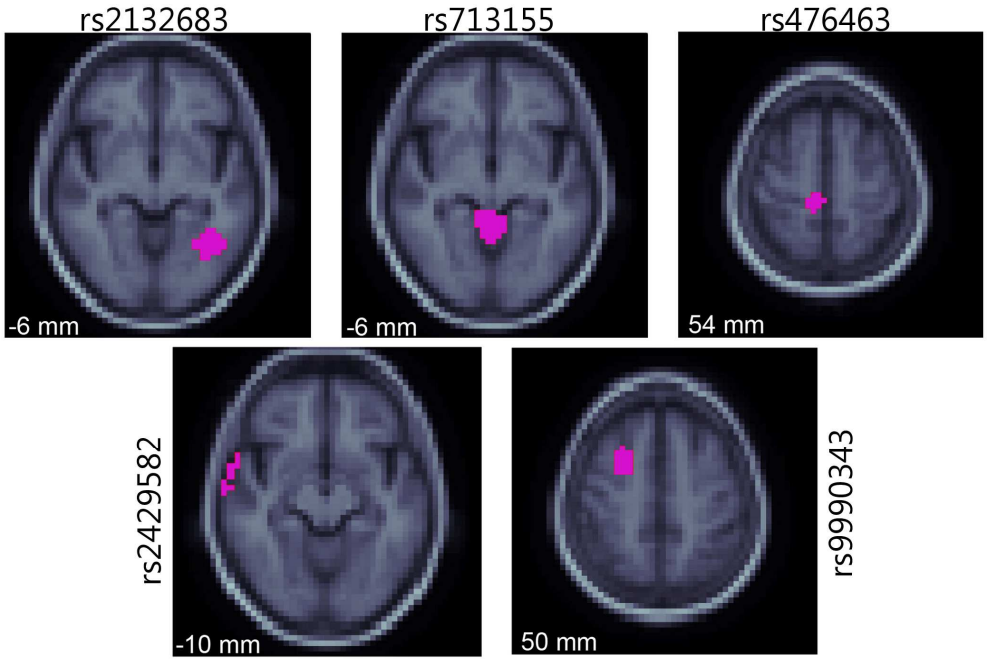
Studying a single imaging measure with a genome-wide search is as limited as picking a single candidate gene from the entire genome – it may not fully reflect how a given genetic variant influences the brain, or it may miss an important effect by being too restrictive. Important links may be overlooked if a gene variant influences a brain feature present but not measured in the images. To broaden the range of measures surveyed in each image, Shen et al. (2010) studied patients with Alzheimer's disease (AD) and mild cognitive impairment (MCI) using whole-brain voxel-based morphometry (VBM; Good et al., 2001) and split the brain into 142 cortical and subcortical ROIs using the segmentation software package FreeSurfer (Fischl et al., 2002). The VBM measure within each region of interest was averaged for each subject and those values were used as traits for GWA scans. One SNP, rs6463843, from the *NXPH1* gene, was significantly associated with grey matter density in the hippocampus, and had broad morphometric effects in a *post-hoc* exploratory analysis. While this study found plausible results, the computation of summaries from regions of interest may miss patterns of effects that lie only partially within the chosen regions of interest. As such, a combination of map-based and ROI-based methods seems ideal.

Some researchers have combined unbiased tests of association across the genome with unbiased searches of the entire brain, instead of relying on summary measures derived from regions of interest. Combining GWA scans with an image-wide search is computationally intensive,

requiring new methods to handle the high dimensionality and multiple statistical comparisons. 3D brain images may contain over 100,000 voxels, and a completely unbiased search may test up to 1 million SNPs for association at each voxel. This is extremely computationally intensive, but can be completed in a feasible time frame if the process is parallelized. Stein et al. (2010b) performed a full GWA scan at each voxel in maps of regional brain volume calculated by tensor-based morphometry (TBM; Leow et al., 2005). 16 billion tests of association were conducted – in a so-called "voxelwise genome-wide association study" (vGWAS). To accommodate the huge number of statistical tests performed, only the most highly associated SNP at each voxel was stored. The *p*-value distribution for the top SNP was modeled as a beta distribution, *Beta*(1, *n*), where *n* is an estimate of the effective number of independent tests performed (Ewens and Grant, 2001). The resulting distribution of minimum *p*-values across the genome, assembled from voxels across the image, was transformed into a uniform distribution in the null case for multiple comparisons correction across the image. FDR was used to correct for multiple comparisons across the image, and to assess whether credible effects had been detected (Benjamini and Hochberg, 1995). Several top SNPs were associated with moderate regional brain volume differences; many were in genes that are expressed in the brain (**Figure 4.3**). However, no SNPs passed the strict correction for multiple comparisons. The Stein et al. study was a *proof of concept*, showing that a completely unbiased search of the genome is feasible with imaging phenotypes. However, the huge correction for multiple comparisons across the image and genome are practically insurmountable unless the effect size or cohort size is very large. In addition, the vGWA study required 27 hours when spread across 500 CPUs; this is more computational power than most researchers typically have access to. Clearly, an optimal balance must be made between pure discovery methods, unconstrained by prior hypotheses, and those

that invoke prior biological information to boost power and reduce the multiple comparisons correction.

Figure 4.3. The 5 most highly associated SNPs identified by vGWAS are shown on slices of an averaged brain MRI template, indicating regions where these SNPs were the most highly associated out of all SNPs (*in purple*). Coordinates refer to the ICBM standard space, and the cohort is the ADNI sample. Adapted from Stein et al., 2010a.



Multivariate imaging genetics methods

Multivariate methods can be used to assess the joint effect of multiple genetic variants simultaneously, and are widely used in genetics (Phillips and Belknap, 2002; Gianola et al., 2002; Cantor et al., 2010). For example, set-based permutation methods use gene annotation information and linkage disequilibrium values to group univariate *p*-values from traditional GWA studies into gene-based test statistics (Hoh et al., 2001; Purcell et al., 2007). Set-based approaches use prior information on gene structure to incorporate all genotyped SNPs in a given

gene into a single test statistic. This can offer, in some cases, greater power than univariate statistical tests to detect SNP effects. Combining univariate *p*-values into a single gene-based test also reduces the total number of tests performed, alleviating the multiple comparisons correction. It can also aggregate the cumulative evidence of association across a gene block to account for allelic heterogeneity (Hoh et al., 2001). Individual SNP *p*-values may not achieve the genome-wide significance level for a traditional GWA study (nominally $p<5x10^{-8}$), but if several SNPs in the same LD block show moderate association, the combined evidence for association may be enough to beat a gene-wide significance level (nominally $p<5x10^{-6}$). For example, one study examined SNPs from the *SORL1* gene for association with hippocampal volume in healthy elderly controls (Bralten et al., 2011). While they did not find evidence for association of individual SNPs in a discovery and replication dataset, a gene-based test found evidence of association in both datasets. Some set-based statistics may be derived from the separate *p*-values from the individual univariate tests, enabling *post-hoc* analysis of published studies. A major issue in applying set-based statistics in imaging genetics is that the permutation procedure applied across SNP groupings would be very computationally intensive. Set-based methods are currently not feasible to apply at >100,000 voxels, as a single gene test takes around 5 min (or 22.8 years to test a single gene at every voxel of the full brain on one CPU). In addition, combining SNPs by *p*-value may miss an important effect where a set of SNPs from the same gene have moderate covariance, but explain different portions of variance in the phenotype. In other words, if they were considered together in the same model, the overall variance explained may be greater than its univariate significance level would imply.

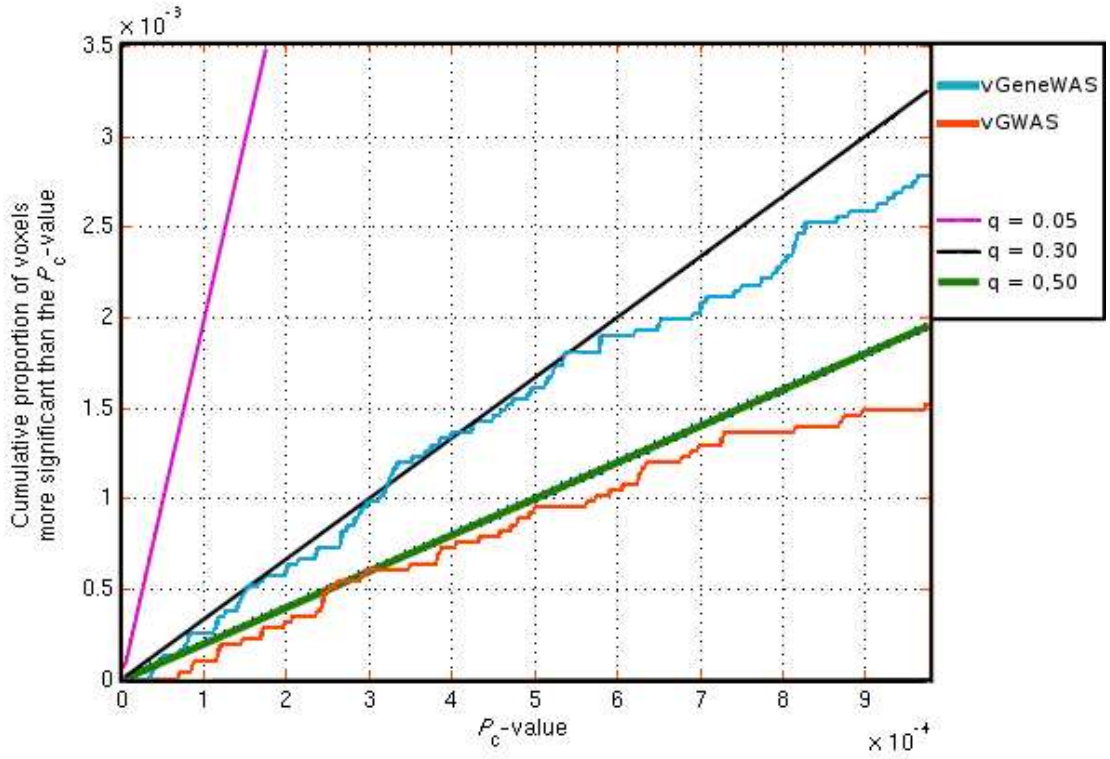
An alternative to set-based methods is to group SNPs into a single statistical model and then test that model for overall association. One classical example of this strategy is multiple linear regression (MLR). However, a problem with applying MLR to genetic data is that SNPs tend to be highly correlated, as they co-segregate in haplotype blocks (Frazer et al., 2007). The MLR is highly sensitive to collinearity among predictors; the inversion step in calculating regression coefficients involves a matrix that is not full rank as the variables are collinear. This leads to wildly inaccurate *Beta* value estimates and standard errors (Kleinbaum, 2007). To avoid collinearity in multivariate analysis of genetic data, dimensionality is often reduced using sparse regression methods, such as penalized or principal components regression.

Some data reduction methods compute a new set of statistically orthogonal variables, for inclusion in a classical MLR model. A data reduction method such as principal components analysis (PCA) transforms a matrix of SNP predictors into a new orthogonal set of predictors, ranked in descending order based on the amount of the variance in the data that each component explains (Jolliffe, 2005). The output of PCA is typically a matrix that explains the same amount of the overall variance as the original predictors, but without the collinearity. As the individual components are sorted by amount of variance they explain, the resulting statistical models can strike an efficient balance between the total variance explained (the number of components to include) and the number of degrees of freedom used (model complexity increases as more variance components are included).

One method, known as *principal components regression* (PCReg) first performs PCA on a set of predictors. It then builds a multiple partial-*F* regression model where the number of components

included is based on the desired proportion of variance to be explained (Massy, 1965). Wang and Abbott (2008) used PCReg to group SNPs into a single multivariate test statistic. Hibar et al. (2011) extended this method to be applicable to images, conducting gene-based tests at each voxel with PCReg. They used an automated method (Altshuler et al., 2005; Hemminger et al., 2006; Hinrichs et al., 2006) to group SNPs based on gene membership, resulting in 18,044 unique genes. Using the set of SNPs in each individual gene as predictors, Hibar et al. used PCReg to assess the degree of association for every gene at every voxel in the full brain. The resulting method was termed a voxelwise "gene-wide" association study (vGeneWAS). By compressing the SNPs into gene-based tests, the total number of tests was reduced to around 500 million tests from the 16 billion tests in vGWAS. However, even with this much smaller number of tests, no genes identified passed correction for multiple comparisons. The most highly associated gene, *GAB2*, showed strong credibility as it is consistently associated with neurodegenerative disorders such as Alzheimer's disease (Reiman et al., 2007). In addition, Hibar et al., (2011) simulated full-brain parametric maps using statistical priors based on their observed data to show that observed clusters of associated genes were larger than would be expected by chance. This provides evidence that vGeneWAS is a valid and powerful multivariate method to detect gene effects in full brain neuroimaging data. A head-to-head comparison of vGWAS and vGeneWAS was also performed on the same datasets. The cumulative distribution function (CDF) plots of *p*-values for each study show that the false discovery rate in the multivariate vGeneWAS was controlled at a lower rate than in the mass univariate vGWAS method (**Figure 4.4**).

Figure 4.4. Cumulative distribution function (CDF) plot of corrected *p*-values from vGeneWAS (Hibar et al., 2011) and vGWAS (Stein et al., 2010) analyses. vGWAS could only be controlled for false-positives at q=0.50 threshold, while vGeneWAS could be controlled for false-positives at q=0.30. The difference in q-value thresholds for the CDF of the *p*-values obtained from both studies on the same dataset suggests that the gene-based analysis is more powerful, though neither study controlled the false-positive rate at the nominal q=0.05 threshold.

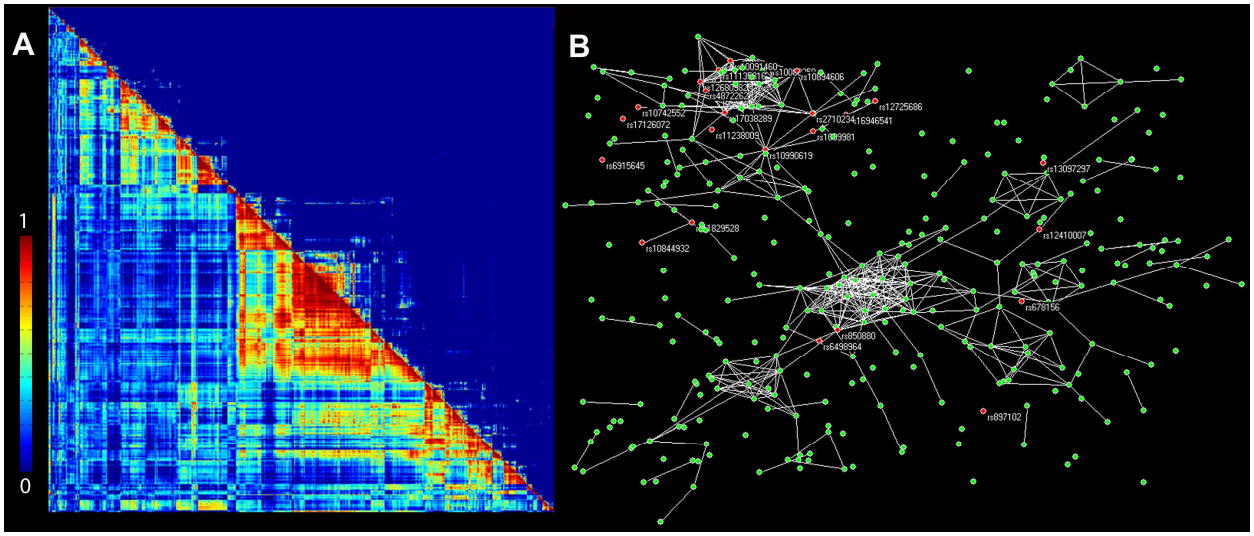


An extension of PCReg and other data reduction techniques is to perform data reduction on *both* the genome and the 3D brain imaging traits. One approach that appears to be promising is parallel independent components analysis (Parallel ICA or PICA; Liu et al., 2009). Parallel ICA works by first performing PCA on a set of SNPs and also a different PCA on a voxelwise imaging measure. Next, a modified version of ICA is applied to both modalities and independent factors from each modality are chosen simultaneously by a correlation measure (hence "parallel" ICA). Selecting imaging features and SNPs together can be more powerful than mass univariate tests of voxelwise imaging traits as the total number of tests is greatly reduced. For example, Lui

et al. (2009) used pre-processed fMRI maps from 43 healthy controls and 20 schizophrenia patients and a preselected set of 384 SNPs chosen for their potential associations with schizophrenia. Via a *t*-test, Lui et al. (2009) demonstrated that genetic components ($p = 0.001$) and fMRI BOLD ($p = 0.0006$) response loadings from parallel ICA were able to distinguish healthy subjects from patients with schizophrenia, with reasonable accuracy. Similar approaches have been applied to structural MRI (Jagannathan et al., 2010). The PICA method is quite promising, but several challenges remain. As Parallel ICA requires an initial round of PCA, it is difficult to recover which SNP sets are contributing to a given component and similarly it is difficult to localize the 3D spatial effect contributing to each component from the image. This may make it difficult to interpret and replicate specific findings. In addition, it is not clear how data reduction methods will perform with whole genome and full brain data. Lui et al. (2009) and Jagannathan et al., (2010) both performed considerable downsampling of the images, reducing the total number of voxels included in the Parallel ICA model. In addition, both studies tested only small sets of pre-selected SNPs instead of data from the full genome, or a standard 500,000 SNP genome-wide scan. The power of Parallel ICA to find common components may be greatly reduced if there is additional noise from genome-wide data. Lui et al. (2009) found that as the amount of random noise increased, so did the number of independent components. As the number of independent components increases, the power to detect associations decreases. Also querying full brain phenotypes for effects of genetic variants, another recently proposed multivariate method by Chiang et al. (2011b), identified patterns of voxels in a DTI image with a common genetic determination, and aggregated them to boost power in GWA (**Figure 4.5**). Approximately 5,000 brain regions were selected, where genetic influences accounted for >60% of the total variation of white matter integrity. From these, a 5,000 x 5,000 correlation matrix

was obtained. Hierarchical clustering was used to select the largest clusters, and these voxels were defined to be ROIs. The mean FAs for these ROIs were then tested for evidence of association with all SNPs genotyped across the genome. By identifying a genetic network that influences white matter integrity over multiple brain regions, Chiang et al. (2011b) were able to boost power to detect associations between FA in these brain areas and SNPs from the whole genome. In all, they identified 24 SNPs with genome-wide significance, which is unusual for a study with fewer than 1000 subjects. To ensure the findings are not false positives, however, simulations of imaging and genomic data may be necessary (as carried out by Vounou et al., see below).

Figure 4.5. Clustering regions of a brain image that have common genetic determination. In a DTI study of twins, the known kinship structure made it possible to estimate the genetic correlation matrix and a "topological overlap (TO)" index matrix. This was used to gauge the similarity of genetic influences on *all pairs of brain regions* (A). The 18 largest clusters – parts of the image with common genetic influences – were selected as regions of interest (ROIs) for GWAS. By associating the mean white matter integrity of these regions with genetic variants, a genetic interconnection network was obtained (B), where each network node represents a single SNP (*colored circles*). The figure shows only those SNPs associated with white matter integrity in at least one ROI with a significance *P*-value $< 10^{-5}$. SNPs whose associations reach genome-wide significance are colored in red, with their names labeled. White lines indicate that SNPs are "connected", i.e., their effects on white matter integrity are strongly correlated. Adapted from Chiang et al., 2010b.

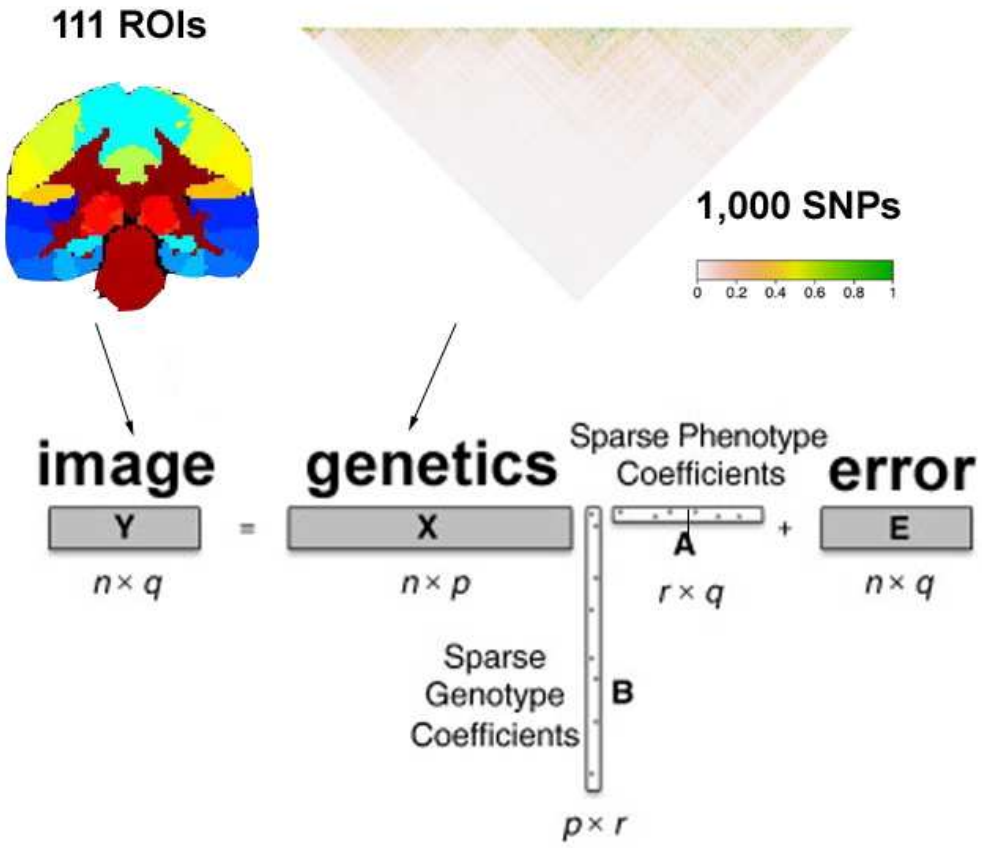


Variants near each other on the genome can be highly correlated due to linkage disequilibrium. This leads to problems if all variants are included in a standard multiple regression model to predict the values of a trait. To address this, many new mathematical methods have been used to handle the high dimensionality in the genome (a $p>>n$ problem) and interactions between genetic variants. These include penalized and sparse regression techniques, such as ridge regression (Hoerl, 1962), the least absolute shrinkage and selection operator (LASSO; Tibshirani, 1996), the elastic net (Zou and Hastie, 2005) and penalized orthogonal-components regression (Zhang et al., 2009; Malo et al, 2008; Cho et al., 2009; Lin et al., 2009; Chen et al., 2010). The various penalty terms (e.g., L^1 in LASSO and L^2 in ridge) in the regularized regression methods can incorporate large numbers of correlated variants with possible interaction terms, in single models. These methods show high statistical power in analyses with both real and simulated data. Although these studies are almost invariably applied to case-control GWA studies, similar approaches may be applied to imaging phenotypes. Kohannim et al. (2011a), for instance, implemented ridge regression to study the association of genomic scanning windows with MRI-derived temporal lobe and hippocampal volume. They reported boosting of power in detecting effects of several SNPs, when compared to univariate imaging GWA. One statistical challenge of such sliding-window approaches is finding optimal window sizes, which can capture the correlation structure in the genomic data without adding excessive degrees of freedom to the model. Kohannim et al. considered several fixed, scanning window sizes (50, 100, 500 and 1000 Kbp) in their study, and found boosting of power in detecting SNPs with different window sizes for different genomic regions. A more flexible approach may incorporate information such as the sample size and variant-specific LD structure into the selection of optimal window sizes for each

genomic region (e.g., Li et al., 2007). This could ensure that SNPs are not missed due to inappropriate window sizes. In addition, L^1-driven methods, such as LASSO, may provide greater detection power by selecting sparse sets of genomic variants in association with imaging measures (Kohannim et al., submitted). As discussed above, however, multivariate methods can be applied not only to the genome, but also to the images, which are also high dimensional and show high spatial correlations. Sparse and penalized models can be useful in these situations as well. In 2010, Vounou et al. applied a sparse reduced-rank regression (sRRR) method to detect whole genome-whole image associations. They computed a matrix of regression coefficients, *C*, whose rank was *p* (number of SNP genotypes) times *q* (the number of imaging phenotypes, or pre-defined anatomical ROIs in their case). They reduced the rank of this large matrix to *r*, by factorizing the matrix into the product of a *p* x *r* matrix, *B*, and an *r* x *q* matrix, *A*, and constraining *A* and *B* to be sparse (**Figure 4.6**). To evaluate the power of their method and compare it to that of mass univariate modeling, Vounou et al. generated realistic, simulated imaging and genetic data. Using the FREGENE (FoRward Evolution of GENomic rEgions) software, and the ADNI baseline T1-weighted MRI dataset, they obtained a simulated dataset, to which they introduced genetic effects in a number of ROIs. It was not feasible for the investigators to consider all possible genetic effect sizes and sample sizes, but they were able to show boosted power for all parameter settings they explored. Setting the desired, reduced rank *r* equal to 2 or 3, they obtained higher sensitivities with sRRR at any given specificity for a sample size of 500. When they increased the sample size from 500 to 1000, they noted gains in sensitivities with sRRR, which were more considerable than the merely linear gains obtained with univariate modeling. They also demonstrated that boosted sensitivities obtained with sRRR increase with higher numbers of SNPs; sensitivity ratios (sRRR / mass univariate modeling)

could be boosted even further to ratios far exceeding 5 (observed with 40,000 SNPs) with numbers of SNPs considered in a typical GWAS (e.g., 500,000 SNPs). Direct power comparisons between association methods on DNA microarray data show that models that incorporate linear combinations of variables perform better than those that perform simple data reduction (Bovelstad et a., 2007). Bovelstad et al., found that the penalized method, ridge regression, was more powerful than LASSO, PCReg, supervised PCReg, and partial least squares regression (PLS), when it comes to predicting survival rates in cancer patients from DNA microarray data. In the future, direct comparisons of methods on imaging genetics data could inform the direction of new methods development.

Figure 4.6. Imaging and genomic data are incorporated into a sparse, reduced-rank regression model, where regions of interest (ROIs) and single nucleotide polymorphisms (SNPs) attain sparse coefficients, simultaneously. This approach can select pertinent SNPs and ROIs. In simulations, it demonstrates higher power than mass univariate models for detecting effects of genetic variants. Adapted with authors' and publishers' permission from Vounou et al. (2010).



Comprehensive modeling of whole-brain voxelwise and genome-wide data remains challenging, due to the high dimensionality of the data. This causes both statistical and computational problems. Recently, there have been new developments applying sparse regression methods to genome-wide data; one such method is *iterative sure independence screening* (ISIS; Fan and Lv, 2009; Fan and Song, 2010; He and Lin, 2011). ISIS is an iterative selection procedure that builds a marginal model using the cyclic coordinate descent (CCD; Friedman et al., 2010) algorithm with the LASSO and combines it with a conditional model of interactions based on pairwise

correlations. The combined model has lower dimensionality, but effects of individual SNPs are still identifiable, as are SNP-SNP interactions. This method appears to be promising for discovery-based searches of the genome. ISIS has not yet been applied to brain images, but it should be feasible. Methods such as ISIS could also be modified to jointly select imaging phenotypes and genomic data as done by Vounou et al., but without first having to select regions of interest or only a small subset of SNPs from the genome.

Conclusion

The field of imaging genetics started with candidate gene studies, where hypotheses about gene action on brain structure and function could be tested in a novel way. More recently, candidate gene studies have been extended to investigate voxelwise associations between genetic variants and images of the brain, to map 3D profiles of genetic effects without requiring *a priori* selection of regions of interest.

To consider the entirety of the genome and discover potentially new variants, however, GWA studies have been introduced to the field of imaging genetics. In these studies, quantitative measures derived from images are considered as intermediate phenotypes, which are in some respects closer to the underlying biology of brain disorders and processes of interest. Despite their unbiased consideration of the whole genome, the standard, univariate GWA approach considers only one SNP at a time and has several limitations. From a genetic perspective, it does not take into account the interdependence between genetic variants due to linkage disequilibrium; and in regard to imaging, such studies typically rely on single summary measures from images, which only weakly represent the wealth of information in a full 3D scan.

Among the most promising applications of imaging genetics are those that use sparse methods to reduce the data dimensionality. Sparse methods create efficient models, and boost power to identify patterns of association. A major advantage of penalized or sparse regression methods is that they accommodate collinearity inherent in the genome and in the images, but they still offer a familiar regression framework to accommodate covariates and confounding variables. Penalized regression models may include a large number of genetic predictors. This may discover genetic effects undetected by other data reduction methods, such as PICA and PCReg. For studies of large 3D statistical maps of imaging phenotypes, methods to penalize the selection of both voxels from the image and associated genetic variants from the genome seem to have higher power than related discovery-based methods. Even so, this is largely an empirical question that depends on the structure of the true signal. Indeed, Vounou et al. demonstrated the increased power of the sRRR method, which favors the selection of an efficient set of regions of interest and a reduced number of SNPs has increased power. A major limitation of penalized methods is that they may fail to converge on a solution when the data dimensions are very high. Even methods designed for $p >> n$ problems such as least angle regression (Effron et al., 2004) tend to fail when given a full 3D imaging phenotype. This illustrates why current implementations of penalized regression in imaging genetics often rely on prior "groupings" of voxels or sliding windows in the genome. These prior groupings do not appear to be motivated by strong prior hypotheses, but by limitations in the statistical modeling. Methods similar to ISIS (Fan and Lv, 2009; Fan and Song, 2010; He and Lin, 2011) designed for ultra-high dimensional datasets will likely be useful for future imaging genetics projects.

Once we have a set of validated genetic variants that affect the brain, multivariate models may be used to combine imaging, genetics, and other physiological biomarkers to predict outcomes in patients with brain disorders. The resulting combination of imaging and genetic data, with other biomarkers, can be used to predict an individual's personalized aggregate risk for specific types of brain disorders. As genomic and proteomic data are added, prognosis and diagnosis may be possible at an earlier stage or more accurate than is possible with current biomarkers. Machine learning algorithms (e.g., decision trees, support vector machines, and neural networks) have shown promise for making disease predictions from genomic and proteomic data (Cruz and Wishart, 2006). Similar approaches may be useful in psychiatry research, and neuroimaging measures such as fiber anisotropy from diffusion imaging may help in making early predictions of brain integrity from genes. In a recent, preliminary study, our group incorporated several candidate polymorphisms in a multi-SNP, machine learning model, to predict personal measures of fiber integrity in the corpus callosum (Kohannim et al., 2011b). Ideally, by incorporating both genomic and proteomic data from larger cohorts, one may be able to obtain personalized "scores" for brain integrity from biomarker profiles. This has considerable implications for prevention and early treatment of brain pathology.

References

Akil H, Brenner S, Kandel E, Kendler KS, King MC, Scolnick E, Watson JD, Zoghbi HY. Medicine. The future of psychiatric research: genomes and neural circuits. Science. 2010; 327(5973):1580-1581.

Altshuler D, Brooks LD, Chakravarti A, Collins FS, Daly MJ, Donnelly P, Gibbs RA, Belmont JW, Boudreau A, Leal SM, Hardenbol P, Pasternak S, Wheeler DA, Willis TD, Yu FL, Yang HM, Zeng CQ, Gao Y, Hu HR, Hu WT, Li CH, Lin W, Liu SQ, Pan H, Tang XL, Wang J, Wang W, Yu J, Zhang B, Zhang QR, Zhao HB, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Shen Y, Yao ZJ, Huang

W, Chu X, He YG, Jin L, Liu YF, Shen YY, Sun WW, Wang HF, Wang Y, Wang Y, Wang Y, Xiong XY, Xu L, Waye MM. Y, Tsui SKW, Xue H, Wong JT. F, Galver IL. M, Fan JB, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier JF, Phillips MS, Roumy S, Sallee C, Verner A, Hudson TJ, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Kwok PY, Cai DM, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller P, Xiao M, Tsui LC, Mak W, Sham PC, Song YQ, Tam PKH, Nakamura Y, Kawaguchi T, Kitamoto T, Morizono T, Nagashima A, Ohnishi Y, Sekine A, Tanaka T, Tsunoda T, Deloukas P, Bird CP, Delgado M, Dermitzakis ET, Gwilliam R, Hunt S, Morrison J, Powell D, Stranger BE, Whittaker P, Bentley DR, Daly MJ, de Bakker PIW, Barrett J, Fry B, Maller J, McCarroll S, Patterson N, Pe'er I, Purcell S, Richter DJ, Sabeti P, Saxena R, Schaffner SF, Varilly P, Stein LD, Krishnan L, Smith AV, Thorisson GA, Chen PE, Cutler DJ, Kashuk CS, Lin S, Abecasis GR, Guan WH, Munro HM, Qin ZHS, Thomas DJ, McVean G, Bottolo L, Eyheramendy S, Freeman C, Marchini J, Myers S, Spencer C, Stephens M, Cardon LR, Clarke G, Evans DM, Morris AP, Weir BS, Tsunoda T, Mullikin JC, Sherry ST, Feolo M, Zhang HC, Zeng CQ, Zhao H, Matsuda I, Fukushima Y, Macer DR, Suda E, Rotimi CN, Adebamowo CA, Ajayi I, Aniagwu T, Marshall PA, Nkwodimmah C, Royal CD. M, Leppert MF, Dixon M, Peiffer A, Qiu RZ, Kent A, Kato K, Niikawa N, Adewole IF, Knoppers BM, Foster MW, Clayton EW, Muzny D, Nazareth L, Sodergren E, Weinstock GM, Wheeler DA, Yakub I, Gabriel SB, Richter DJ, Ziaugra L, Birren BW, Wilson RK, Fulton LL, Rogers J, Burton J, Carter NP, Clee CM, Griffiths M, Jones MC, McLay K, Plumb RW, Ross MT, Sims SK, Willey DL, Chen Z, Han H, Kang L, Godbout M, Wallenburg JC, Archeveque PL, Bellemare G, Saeki K, Wang HG, An DC, Fu HB, Li Q, Wang Z, Wang RW, Holden AL, Brooks LD, McEwen JE, Bird CR, Guyer MS, Nailer PJ, Wang VO, Peterson JL, Shi M, Spiegel J, Sung LM, Witonsky J, Zacharia LF, Kennedy K, Jamieson R, Stewart J, the International HapMap Consortium. A haplotype map of the human genome. Nature. 2005; 437(7063):1299-1320.

Beaulieu C. The basis of anisotropic water diffusion in the nervous system - a technical review. NMR Biomed. 2002; 15(7-8):433-455.

Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol. 1995; 57(1):289-300.

Bovelstad HM, Nygard S, Storvold HL, Aldrin M, Borgan O, Frigessi A, Lingjaerde OC. Predicting survival from microarray data - a comparative study. Bioinformatics. 2007; 23(16):2080-2087.

Bralten J, Arias-Vasquez A, Makkinje R, Veltman JA, Brunner HG, Fernandez G, Rijpkema M, Franke B. Association of the Alzheimer's gene SORL1 with hippocampal volume in young, healthy adults. Am J Psychiatry. 2011; 168(10):1083-1089.

Braskie MN, Jahanshad N, Stein JL, Barysheva M, McMahon KL, de Zubicaray GI, Martin NG, Wright MJ, Ringman JM, Toga AW, Thompson PM. Common Alzheimer's disease risk variant within the CLU gene affects white matter microstructure in young adults. J Neurosci. 2011(a); 31(18):6764-6770.

Braskie MN, Ringman JM, Thompson PM. Neuroimaging measures as endophenotypes in Alzheimer's disease. Int J Alzheimers Dis. 2011(b); 31:490140.

Cantor RM, Lange K, Sinsheimer JS. Prioritizing GWAS results: a review of statistical methods and recommendations for their application. Am J Hum Genet. 2010; 86(1):6-22.

Chen LS, Hutter CM, Potter JD, Liu Y, Prentice RL, Peters U, Hsu L. Insights into colon cancer etiology via a regularized approach to gene set analysis of GWAS data. Am J Hum Genet. 2010; 86(6):860-871.

Chiang MC, Barysheva M, Shattuck DW, Lee AD, Madsen SK, Avedissian C, Klunder AD, Toga AW, McMahon KL, de Zubicaray GI, Wright MJ, Srivastava A, Balov N, Thompson PM. Genetics of brain fiber architecture and intellectual performance. J Neurosci. 2009; 29(7):2212-2224.

Chiang MC, Barysheva M, Toga AW, Medland SE, Hansell NK, James MR, McMahon KL, de Zubicaray GI, Martin NG, Wright MJ, Thompson PM. BDNF gene effects on brain circuitry replicated in 455 twins. NeuroImage. 2011a; 55(2):448-454.

Chiang MC, Barysheva M, McMahon KL, de Zubicaray GI, Johnson K, Martin NG, Toga AW, Wright MJ, Thompson PM. Hierarchical clustering of the genetic connectivity matrix reveals the network topology of gene action on brain microstructure: an N=531 twin study. International Workshop on Biomedical Imaging. Chicago, IL, USA. 2011b.

Cho S, Kim H, Oh S, Kim K, Park T. Elastic-net regularization approaches for genome-wide association studies of rheumatoid arthritis. BMC Proc. 2009; 15(3Suppl):S7-S25.

Chumbley J, Worsley K, Flandin G, Friston K. Topological FDR for neuroimaging. NeuroImage. 2010; 49(4):3057-3064.

Cruz JA, Wishart DS. Applications of machine learning in cancer prediction and prognosis. Cancer Inform. 2007; 2:59-77.

de Bakker PI, Ferreira MA, Jia X, Neale BM, Raychaudhuri S, Voight BF. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. Hum Mol Genet. 2008; 17(R2):R122-R128.

Efron B, Hastie T, Johnstone I, Tibshirani R. Least angle regression. Ann Stat. 2004; 32(2):407-451.

Ewens WJ, Grant G. Statistical Methods in Bioinformatics: An Introduction. Springer, New York. 2001.

Fan JQ, Lv JC. Sure independence screening for ultrahigh dimensional feature space. J R Stat Soc Series B Stat Methodol. 2008; 70(5):849-883.

Fan JQ, Song R. Sure independence screening in generalized linear models with Np-dimensionality. Ann Stat. 2010; 38(6):3567-3604.

Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, Van Der Kouwe A, Killiany R, Kennedy D, Klaveness S, Montillo A, Makris N, Rosen B, Dale AM. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron. 2002; 33(3):341-355.

Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, Perry JR, Elliott KS, Lango H, Rayner NW, Shields B, Harries LW, Barrett JC, Ellard S, Groves CJ, Knight B, Patch AM, Ness AR, Ebrahim S, Lawlor DA, Ring SM, Ben-Shlomo Y, Jarvelin MR, Sovio U, Bennett AJ, Melzer D, Ferrucci L, Loos RJ, Barroso I, Wareham NJ, Karpe F, Owen KR, Cardon LR, Walker M, Hitman GA, Palmer CN, Doney AS, Morris AD, Smith GD, Hattersley AT, McCarthy MI. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science. 2007; 316(5826):889-894.

Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang

Q, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Altshuler D, Shen Y, Yao Z, Huang W, Chu X, He Y, Jin L, Liu Y, Sun W, Wang H, Wang Y, Xiong X, Xu L, Waye MM, Tsui SK, Xue H, Wong JT, Galver LM, Fan JB, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier JF, Phillips MS, Roumy S, Sallee C, Verner A, Hudson TJ, Kwok PY, Cai D, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller P, Xiao M, Tsui LC, Mak W, Song YQ, Tam PK, Nakamura Y, Kawaguchi T, Kitamoto T, Morizono T, Nagashima A, Ohnishi Y, Sekine A, Tanaka T, Tsunoda T, Deloukas P, Bird CP, Delgado M, Dermitzakis ET, Gwilliam R, Hunt S, Morrison J, Powell D, Stranger BE, Whittaker P, Bentley DR, Daly MJ, de Bakker PI, Barrett J, Chretien YR, Maller J, McCarroll S, Patterson N, Pe'er I, Price A, Purcell S, Richter DJ, Sabeti P, Saxena R, Schaffner SF, Sham PC, Varilly P, Altshuler D, Stein LD, Krishnan L, Smith AV, Tello-Ruiz MK, Thorisson GA, Chakravarti A, Chen PE, Cutler DJ, Kashuk CS, Lin S, Abecasis GR, Guan W, Li Y, Munro HM, Qin ZS, Thomas DJ, McVean G, Auton A, Bottolo L, Cardin N, Eyheramendy S, Freeman C, Marchini J, Myers S, Spencer C, Stephens M, Donnelly P, Cardon LR, Clarke G, Evans DM, Morris AP, Weir BS, Tsunoda T, Mullikin JC, Sherry ST, Feolo M, Skol A, Zhang H, Zeng C, Zhao H, Matsuda I, Fukushima Y, Macer DR, Suda E, Rotimi CN, Adebamowo CA, Ajayi I, Aniagwu T, Marshall PA, Nkwodimmah C, Royal CD, Leppert MF, Dixon M, Peiffer A, Qiu R, Kent A, Kato K, Niikawa N, Adewole IF, Knoppers BM, Foster MW, Clayton EW, Watkin J, Gibbs RA, Belmont JW, Muzny D, Nazareth L, Sodergren E, Weinstock GM, Wheeler DA, Yakub I, Gabriel SB, Onofrio RC, Richter DJ, Ziaugra L, Birren BW, Daly MJ, Altshuler D, Wilson RK, Fulton LL, Rogers J, Burton J, Carter NP, Clee CM, Griffiths M, Jones MC, McLay K, Plumb RW, Ross MT, Sims SK, Willey DL, Chen Z, Han H, Kang L, Godbout M, Wallenburg JC, L'Archevêque P, Bellemare G, Saeki K, Wang H, An D, Fu H, Li Q, Wang Z, Wang R, Holden AL, Brooks LD, McEwen JE, Guyer MS, Wang VO, Peterson JL, Shi M, Spiegel J, Sung LM, Zacharia LF, Collins FS, Kennedy K, Jamieson R, Stewart J. A second generation human haplotype map of over 3.1 million SNPs. Nature. 2007; 449(7164):851-861.

Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw. 2010: 33(1):1-22.

Gianola D, Perez-Enciso M, Toro MA. On marker-assisted prediction of genetic value: beyond the ridge. Genetics. 2003; 163(1):347-365.

Glahn DC, Thompson PM, Blangero J. Neuroimaging endophenotypes: strategies for finding genes influencing brain structure and function. Hum Brain Mapp. 2007; 28(6):488-501.

Good CD, Johnsrude IS, Ashburner J, Henson RN, Friston KJ, Frackowiak RS. A voxel-based morphometric study of ageing in 465 normal adult human brains. NeuroImage. 2001; 14(1):21-36.

Gottesman II, Gould TD. The endophenotype concept in psychiatry: etymology and strategic intentions. Am J Psychiatry. 2003; 160(4):636-645.

Hall MH, Smoller JW. A new role for endophenotypes in the GWAS era: functional characterization of risk variants. Harv Rev Psychiatry. 2010; 18(1):67-74.

Hariri AR, Mattay VS, Tessitore A, Kolachana B, Fera F, Goldman D, Egan MF, Weinberger DR. Serotonin transporter genetic variation and the response of the human amygdala. Science. 2002; 297(5580):400-403.

He Q, Lin DY. A variable selection method for genome-wide association studies. Bioinformatics. 2011; 27(1):1-8.
Heinz A, Jones DW, Mazzanti C, Goldman D, Ragan P, Hommer D, Linnoila M, Weinberger DR. A relationship between serotonin transporter genotype and in vivo protein expression and alcohol neurotoxicity. Biol Psychiatry. 2000: 47(7):643-649.
Hemminger BM, Saelim B, Sullivan PF. TAMAL: an integrated approach to choosing SNPs for genetic studies of human complex traits. Bioinformatics. 2006; 22(5):626-627.
Hibar DP, Stein JL, Kohannim O, Jahanshad N, Saykin AJ, Shen L, Kim S, Pankratz N, Foroud T, Huentelman MJ, Potkin SG, Jack CR Jr, Weiner MW, Toga AW, Thompson PM. Voxelwise gene-wide association study (vGeneWAS): multivariate gene-based association testing in 731 elderly subjects. NeuroImage. 2011; 56(4):1875-1891.
Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci USA. 2009; 106(23):9362-9367.
Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H, Diekhans M, Furey TS, Harte RA, Hsu F, Hillman-Jackson J, Kuhn RM, Pedersen JS, Pohl A, Raney BJ, Rosenbloom KR, Siepel A, Smith KE, Sugnet CW, Sultan-Qurraie A, Thomas DJ, Trumbower H, Weber RJ, Weirauch M, Zweig AS, Haussler D, Kent WJ. The UCSC genome browser database: update 2006. Nucleic Acids Res. 2006; 34:D590-D598.
Ho AJ, Stein JL, Hua X, Lee S, Hibar DP, Leow AD, Dinov ID, Toga AW, Saykin AJ, Shen L, Foroud T, Pankratz N, Huentelman MJ, Craig DW, Gerber JD, Allen AN, Corneveaux JJ, Stephan DA, DeCarli CS, DeChairo BM, Potkin SG, Jack CR Jr, Weiner MW, Raji CA, Lopez OL, Becker JT, Carmichael OT, Thompson PM, Alzheimer's Disease Neuroimaging Initiative. A commonly carried allele of the obesity-related FTO gene is associated with reduced brain volume in the healthy elderly. Proc Natl Acad Sci USA. 2010; 107(18):8404-8409.
Hoerl AE. Application of ridge analysis to regression problems. Chem Eng Prog. 1962; 58, 54-59.
Hoh J, Wille A, Ott J. Trimming, weighting, and grouping SNPs in human case-control association studies. Genome Res. 2001; 11(12):2115-2119.
Ioannidis JP. Why most published research findings are false. PLoS Med. 2005; 2(8):e124.
Jagannathan K, Calhoun VD, Gelernter J, Stevens MC, Liu J, Bolognani F, Windemuth A, Ruaño G, Assaf M, Pearlson GD. Genetic associations of brain structural networks in schizophrenia: a preliminary study. Biol Psychiatry. 2010; 68(7):657-666.
Jolliffe IT. Principal Component Analysis. Springer, New York. 2002.
Joyner AH, Roddey JC, Bloss CS, Bakken TE, Rimol LM, Melle I, Agartz I, Djurovic S, Topol EJ, Schork NJ, Andreassen OA, Dale AM. A common MECP2 haplotype associates with reduced cortical surface area in humans in two independent populations. Proc Natl Acad Sci USA. 2009; 106(36):15483-15488.
Kleinbaum DG. Applied Regression Analysis and Other Multivariable Methods. Belmont, CA: Brooks/Cole. 2007.
Kochunov P, Glahn DC, Lancaster JL, Winkler AM, Smith S, Thompson PM, Almasy L, Duggirala R, Fox PT, Blangero J. Genetics of microstructure of cerebral white matter using diffusion tensor imaging. NeuroImage 2010; 53(3):1109-1116.
Kohannim O, Hibar DP, Stein JL, Jahanshad N, Jack CR Jr, Weiner MW, Toga AW, Thompson PM. Boosting power to detect genetic associations in imaging using multi-locus, genome-

wide scans and ridge regression. International Workshop on Biomedical Imaging (ISBI), Chicago, IL, USA. 2011a.
Kohannim O, Jahanshad N, Braskie MN, Stein JL, Chiang MC, Reese AH, Toga AW, McMahon KL, de Zubicaray GI, Medland SE, Montgomery GM, Whitfield JB, Martin NG, Wright MW, Thompson PM. Personalized prediction of brain fiber integrity in 396 young adults based on genotyping of multiple common genetic variants. Society for Neuroscience. Washington, DC, USA. 2011b.
Leow A, Huang SC, Geng A, Becker J, Davis S, Toga A, Thompson P. Inverse consistent mapping in 3D deformable image registration: its construction and statistical properties. Inf Process Med Imaging. 2005; 19():493-503.
Li Y, Sung W, Liu JJ. Association mapping via regularized regression analysis of single-nucleotide-polymorphism haplotypes in variable-sized sliding windows. Am J Hum Genet. 2007; 80(4):705-715.
Lin Y, Zhang M, Wang L, Pungpapong V, Fleet JC, Zhang D. Simultaneous genome-wide association studies of anti-cyclic citrullinated peptide in rheumatoid arthritis using penalized orthogonal-components regression. BMC Proc. 2009; 15(3Suppl):S7-S20.
Liu J, Pearlson G, Windemuth A, Ruano G, Perrone-Bizzozero NI, Calhoun V. Combining fMRI and SNP data to investigate connections between brain function and genetics using parallel ICA. Hum Brain Mapp. 2009; 30(1):241-255.
Malo N, Libiger O, Schork NJ. Accommodating linkage disequilibrium in genetic-association analyses via ridge regression. Am J Hum Genet. 2008; 82(2):375-385.
Massy WF. Principal components regression in exploratory statistical research. J Am Stat Assoc. 1965; 60, 234-256.
McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, Hirschhorn JN. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet. 2008; 9(5):356-369.
Meyer-Lindenberg A, Weinberger DR. Intermediate phenotypes and genetic mechanisms of psychiatric disorders. Nat Rev Neurosci. 2006; 7(10):818-827.
Munafo MR, Brown SM, Hariri AR. Serotonin transporter (5-HTTLPR) genotype and amygdala activation: a meta-analysis. Biol Psychiatry. 2008; 63(9):852-857.
Pearson TA, Manolio TA. How to interpret a genome-wide association study. JAMA. 2008; 299(11):1335-1344.
Phillips TJ, Belknap JK. Complex-trait genetics: emergence of multivariate strategies. Nat Rev Neurosci. 2002; 3(6):478-485.
Potkin SG, Guffanti G, Lakatos A, Turner JA, Kruggel F, Fallon JH, Saykin AJ, Orro A, Lupoli S, Salvi E, Weiner M, Macciardi F, Alzheimer's Disease Neuroimaging Initiative. Hippocampal atrophy as a quantitative trait in a genome-wide association study identifying novel susceptibility genes for Alzheimer's disease. PLoS One. 2009a; 4(8):e6501.
Potkin SG, Turner JA, Fallon JA, Lakatos A, Keator DB, Guffanti G, Macciardi F. Gene discovery through imaging genetics: identification of two novel genes associated with schizophrenia. Mol Psychiatry. 2009b; 14(4):416-428.
Psychiatric GWAS Consortium Coordinating Committee. Cichon S, Craddock N, Daly M, Faraone SV, Gejman PV, Kelsoe J, Lehner T, Levinson DF, Moran A, Sklar P, Sullivan PF. Genomewide association studies: history, rationale, and prospects for psychiatric disorders. Am J Psychiatry. 2009; 166(5):540-556.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, De Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81(3):559-575.

Reiman EM, Webster JA, Myers AJ, Hardy J, Dunckley T, Zismann VL, Joshipura KD, Pearson JV, Hu-Lince D, Huentelman MJ, Craig DW, Coon KD, Liang WS, Herbert RH, Beach T, Rohrer KC, Zhao AS, Leung D, Bryden L, Marlowe L, Kaleem M, Mastroeni D, Grover A, Heward CB, Ravid R, Rogers J, Hutton ML, Melquist S, Petersen RC, Alexander GE, Caselli RJ, Kukull W, Papassotiropoulos A, Stephan DA. GAB2 alleles modify Alzheimer's risk in APOE epsilon4 carriers. Neuron. 2007; 54(5):713-720.

Shaw LM, Korecka M, Clark CM, Lee VM, Trojanowski JQ. Biomarkers of neurodegeneration for diagnosis and monitoring therapeutics. Nat Rev Drug Discov. 2007; 6(4):295-303.

Shen L, Kim S, Risacher SL, Nho K, Swaminathan S, West JD, Foroud T, Pankratz N, Moore JH, Sloan CD, Huentelman MJ, Craig DW, Dechairo BM, Potkin SG, Jack CR Jr, Weiner MW, Saykin AJ. Whole genome association study of brain-wide imaging phenotypes for identifying quantitative trait loci in MCI and AD: a study of the ADNI cohort. NeuroImage. 2010; 53(3):1051-1063.

Stein JL, Hibar DP, Madsen SK, Khamis M, Mcmahon KL, De Zubicaray GI, Hansell NK, Montgomery GW, Martin NG, Wright MJ, Saykin AJ, Jack CR Jr, Weiner MW, Toga AW, Thompson PM. Discovery and replication of dopamine-related gene effects on caudate volume in young and elderly populations (N=1198) using genome-wide search. Mol Psychiatry. 2011; 16(9):927-937.

Stein JL, Hua X, Morra JH, Lee S, Hibar DP, Ho AJ, Leow AD, Toga AW, Sul JH, Kang HM, Eskin E, Saykin AJ, Shen L, Foroud T, Pankratz N, Huentelman MJ, Craig DW, Gerber JD, Allen AN, Corneveaux JJ, Stephan DA, Webster J, DeChairo BM, Potkin SG, Jack CR Jr, Weiner MW, Thompson PM, Alzheimer's Disease Neuroimaging Initiative. Genome-wide analysis reveals novel genes influencing temporal lobe structure with relevance to neurodegeneration in Alzheimer's disease. NeuroImage. 2010a; 51(2):542-554.

Stein JL, Hua X, Lee S, Ho AJ, Leow AD, Toga AW, Saykin AJ, Shen L, Foroud T, Pankratz N, Huentelman MJ, Craig DW, Gerber JD, Allen AN, Corneveaux JJ, Dechairo BM, Potkin SG, Weiner MW, Thompson P. Voxelwise genome-wide association study (vGWAS). NeuroImage. 2010b; 53(3):1160-1174.

The ENIGMA Consortium. Genome-wide association meta-analysis of hippocampal volume: results from the ENIGMA Consortium. Organization for Human Brain Mapping Conference. Quebec City, Canada. 2011.

Thomason ME, Dougherty RF, Colich NL, Perry LM, Rykhlevskaia EI, Louro HM, Hallmayer JF, Waugh CE, Bammer R, Glover GH, Gotlib IH. COMT genotype affects prefrontal white matter pathways in children and adolescents. NeuroImage. 2010; 53(3):926-934.

Thomason ME, Thompson PM. Diffusion imaging, white matter, and psychopathology. Annu Rev Clin Psychol. 2011; 7:63-85.

Thompson PM, Cannon TD, Narr KL, Van Erp T, Poutanen VP, Huttunen M, Lonnqvist J, Standertskjold-Nordenstam CG, Kaprio J, Khaledy M, Dail R, Zoumalan CI, Toga AW. Genetic influences on brain structure. Nat Neurosci. 2001; 4(12):1253-1258.

Tibshirani R. Regression Shrinkage and Selection via the Lasso. J R Stat Soc Series B Stat Methodol. 1996; 58(1):267-288.

Vounou M, Nichols TE, Montana G, Alzheimer's Disease Neuroimaging Initiative. Discovering genetic associations with high-dimensional neuroimaging phenotypes: a sparse reduced-rank regression approach. NeuroImage. 2010; 53(3):1147-1159.

Wang K, Abbott D. A principal components regression approach to multilocus genetic association studies. Genet Epidemiol. 32(2):108-118.

Winkler AM, Kochunov P, Blangero J, Almasy L, Zilles K, Fox PT, Duggirala R, Glahn DC. Cortical thickness or grey matter volume? The importance of selecting the phenotype for imaging genetics studies. NeuroImage 2010; 53(3):1135-1146.

Yang J, Benyamin B, Mcevoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, Goddard ME, Visscher PM. Common SNPs explain a large proportion of the heritability for human height. Nat Genet. 2010; 42(7):565-569.

Zeggini E, Ioannidis JP. Meta-analysis in genome-wide association studies. Pharmacogenomics. 2009; 10(2):191-201.

Zhang D, Lin Y, Zhang M. Penalized orthogonal-components regression for large p small n Data. Electron J Stat. 3:781-796.

Zou H, Hastie T. Regularization and variable selection via the elastic net. J R Stat Soc Series B Stat Methodol. 2005; 67(2):301-320.

Chapter 5. Multilocus genome-wide scan with ridge regression for gene discovery

Introduction

Imaging genetics is a new, emerging field in biomedical imaging. It aims to discover specific genetic variants, such as SNPs, that account for differences in anatomy and function. Analysis of image databases may pick up gene effects more efficiently than analysis of clinical or behavioral test scores, as imaging measures have high precision and reproducibility. Many image-derived measures such as hippocampal and caudate volumes are highly heritable and may be more directly influenced by genetic variation (Thompson et al., 2010). Already, genome-wide association studies (GWAS) of large MRI datasets (e.g., ADNI; N=740 subjects) have identified new candidate Alzheimer's Disease (AD) risk genes including *GRIN2B* (Stein et al., 2010) and *TOMM40* (Potkin et al., 2009) that are associated with regional brain volumes.

Several international efforts, such as the Enigma project (http://enigma.loni.ucla.edu) are currently searching for genetic variants that affect brain structure and function, using databases of up to 10,000 images (The ENIGMA Consortium, 2011). Imaging GWAS studies, so far, only consider the independent effect of each variation in the genome. This ignores useful information from multiple SNPs in the same gene, and across the genome. Many brain measures are heritable; each SNP has a weak effect on its own, but moderate to strong effects are likely when all SNP effects are aggregated across the whole genome. To model effects of large numbers of predictors with weak effects, machine learning approaches, including penalized regression, artificial neural networks, support vector machines, and adaptive boosting methods, have been introduced for GWAS in the last few years. These so-called *multi-locus* genetic methods model

the combined effect of large numbers of SNPs, to explain more of the genetic contribution to particular phenotypes (Szymczak et al., 2009). These approaches have been applied to several phenotypes related to rheumatoid arthritis (Sun et al., 2009) and coronary heart disease (Kim et al., 2009), but they have not yet been applied to the analysis of brain image databases.

Ridge regression (Hoerl, 1959; Hoerl, 1962; Hoerl and Kennard, 1970) is one of several penalized regression methods for high-dimensional data analysis. Ridge regression works in many situations where ordinary multiple regression breaks down. It handles large numbers of highly correlated predictors, such as SNPs. Malo et al. (2008) showed that ridge regression tends to outperform standard, univariate regression in GWAS studies, except when only one single SNP affects the measures of interest. Ridge regression has been previously applied to the study of multiple SNPs (Malo et al., 2008; Sun et al., 2009), but not at a genome-wide level, and not in the field of imaging. Here, we apply this modified multiple regression approach to a genome-wide analysis of the baseline brain MRI data from ADNI. We set out to find SNPs associated with measures of temporal lobe volume, and hippocampal volume, based on structural MRI. We hypothesize that our novel, genome-wide, penalized regression GWAS approach would help identify new candidate SNPs associated with imaging measures, including SNPs missed using standard univariate GWAS, which tests them independently.

Methods

A. Structural MRI Measures

All subjects were scanned with a standard MRI protocol developed for ADNI. Hippocampal volumes were generated by an automatic segmentation method developed by our group, based on

adaptive boosting (Morra et al., 2008). Temporal lobe volumes were derived from an anatomically defined region-of-interest (ROI) on three-dimensional atrophy maps generated with tensor-based morphometry (TBM), a well-established method for mapping volumetric differences in the brain (Hua et al., 2008). Data were available for 740 ADNI subjects (173 AD, 361 MCI, 206 controls; 438 men/302 women; mean ± SD age: 75.55 ± 6.79 years). MRI-derived measures of hippocampal volume were computed for a subset of 696 ADNI subjects (162 AD, 343 MCI, 191 controls; 405 men/ 291 women; mean ± SD age: 75.36 ± 6.77 years). These measures were adjusted for sex and age.

B. Genotypes

Genotyping procedures for ADNI are thoroughly described in Saykin et al. (2010). As described in Stein et al. (2010b), genotypes were imputed to remove missing information and to compute the effective number of statistical tests across the genome; we also extracted SNPs that had minor allele frequencies greater than 10%, and Hardy-Weinberg equilibrium *p*-values more strict than 5.7×10^{-7}.

C. Ridge Regression

Hoerl introduced ridge regression as a variant of multiple regression (Hoerl, 1959; Hoerl, 1962; Hoerl and Kennard, 1970). It is designed to handle high-dimensional data, in cases where high correlations among the predictors would lead standard multiple regression methods to fail. Several variants of the same method were independently discovered in separate branches of mathematics and statistics. Tikhonov regularization (Tikhonov and Arsenin, 1977) is a related

concept, for solving inverse problems. It enforces solutions to be smooth, by minimizing a penalty function that controls the regularity of the solution.

In standard multiple regression, coefficients, β_i, are obtained by minimizing the residual sum of squares of the data, after fitting the regression model, which yields:

$$\beta = (X^t X)^{-1} X^t Y \qquad (1)$$

Here X is an n x p matrix of p predictors or SNPs and Y is an n-dimensional vector of imaging measures obtained from n subjects. In analyses such as ours, where there are many highly correlated predictors, standard multiple regression fails because the $X^t X$ matrix is highly ill-conditioned or not invertible.

Ridge regression addresses this by introducing a positive *shrinkage* parameter, λ, to obtain regression coefficients as follows:

$$\beta = (X^t X + \lambda I)^{-1} X^t Y \qquad (2)$$

Here I is the p x p identity matrix. As its name suggests, λ constrains the size of the regression coefficients by shrinking their variance to a specific, tunable extent. The idea of solving a regression equation with coefficients that are as small as possible (and forcing some to zero) is highly related to *compressed sensing* in computer vision (Donoho, 2006) or "L1-norm" minimization methods in mathematics, such as Bregman splitting (Goldstein and Osher, 2008). Coefficients are standardized after dividing them by their standard errors, which are the square roots of the diagonal elements of the variance-covariance matrix:

$$\mathrm{var}(\beta) = (X^t X + \lambda I)^{-1} X^t X (X^t X + \lambda I)^{-1} \sigma^2 \quad (3)$$

(Chatterjee and Price, 1977). Although we implemented this more traditional form of the variance-covariance matrix for our ridge regression analyses, other studies have considered alternative forms of this matrix, such as:

$$\text{var}(\beta) = \sigma^2 (X^t X + \lambda I)^{-1} \qquad (4)$$

(Halawa and Bassiouni, 2000). *P*-values are then obtained from the *t*-distributed standardized coefficients, using the following formula to compute the effective number of degrees of freedom (EDF), formulated specifically for ridge regression (Malo et al., 2008):

$$EDF = trace(X(X^t X + \lambda I)^{-1} X^t) \qquad (5)$$

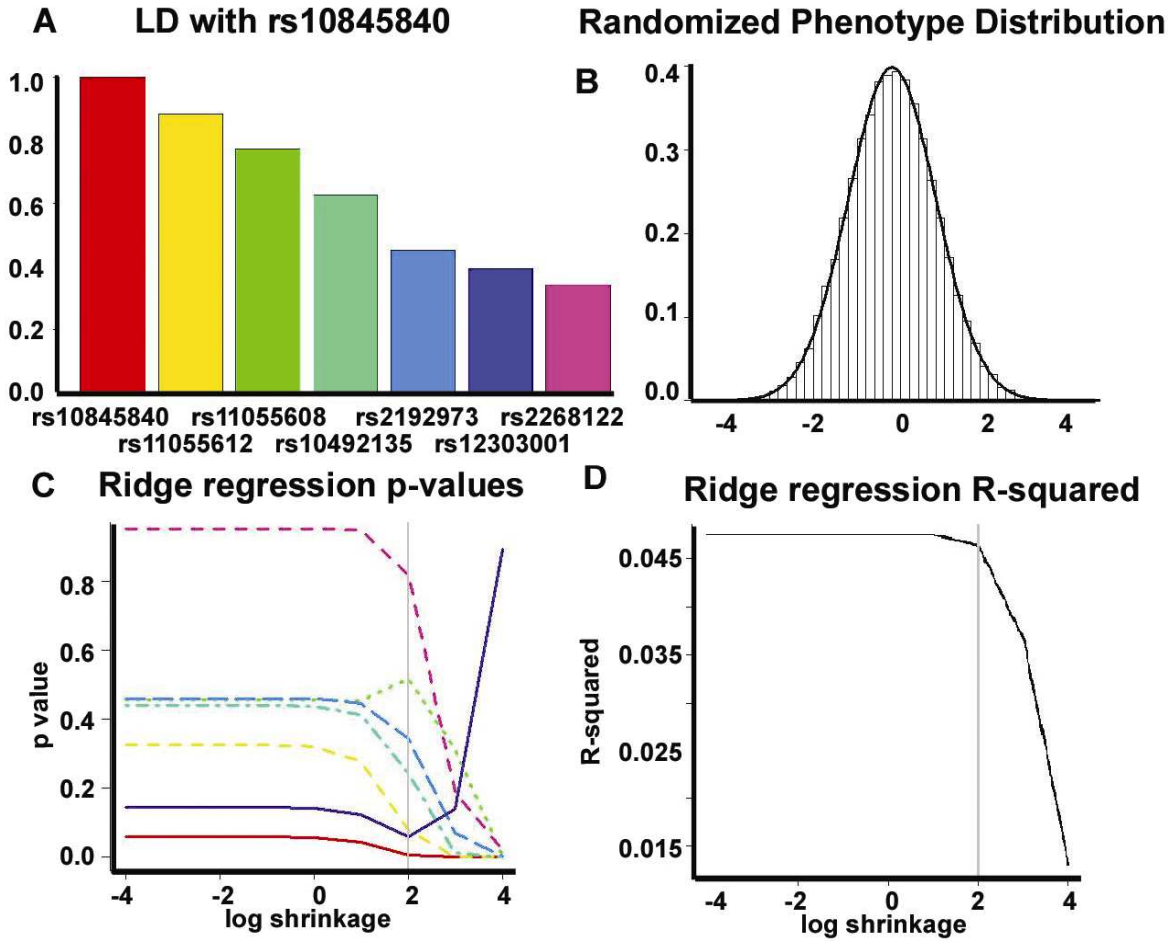
As the ridge regression shrinkage parameter, λ, gets closer to zero, the model behaves more like multiple regression with a similar coefficient of determination (R^2). As λ approaches infinity, the model acts more similarly to univariate regression and the R^2 or predictability of the model decreases. There is a need to find the optimal shrinkage parameter, offering sufficient shrinkage to allow for multi-collinearity, but not so high that the model loses its predictive ability. Several statistical methods exist to select the best shrinkage parameter for a ridge regression model. These include the Hoerl, Kennard and Baldwin estimator (Hoerl et al., 1975), and the Lawless and Wang estimator (Lawless and Wang, 1976). Here, we base our shrinkage estimation on the latter estimator (*LW* estimator, below), which, through a Bayesian approach, estimates λ as follows:

$$\lambda = p\sigma^2 / \sum_i \varepsilon_i (Q^t \beta)_i^2 \qquad (6)$$

Here, the ε_i are eigenvalues of $X^t X$, Q is a p x p matrix, with the eigenvectors of $X^t X$ as column vectors, and β represents the vector of standard multiple regression coefficients.

To illustrate the concept of ridge regression and test how the *LW* estimator performs with our data, we considered six SNPs in various incrementally decreasing levels of linkage disequilibrium, or correlation, with rs10845840, a SNP in the *GRIN2B* gene that our team previously identified with standard GWAS to be associated with temporal lobe volume (Stein et al., 2010). We investigated how the ridge regression shrinkage parameter affected the power to detect the effect of the main SNP of interest, when other adjacent SNPs were added to the model. By randomly permuting the imaging data 10,000 times (i.e., assigning the images to the wrong subjects), we ensured that the reference distribution of *t* statistics was appropriate for assessing significance (results are shown in **Figure 5.1**).

Figure 5.1. (a) The degrees of statistical correlation, across subjects, of the selected *GRIN2B* SNPs are shown with respect to rs10845840. In genetics, this is known as linkage disequilibrium (LD). (b) shows the distribution (*histogram*) of standardized ridge regression coefficients with shuffled phenotypes, along with a *t*-distribution with the same effective number of degrees of freedom (*curve*). The point of this experiment is to show that when the data are null, by construction, the *t*-statistic computed from the ridge regression formula correctly controls for false positive associations in null data. (c) Ridge regression *P*-values are graphed against the level of shrinkage. Each color represents a SNP, corresponding to panel (a). The gray line represents the optimal shrinkage determined by the *LW* estimator. (d) The regression model's R^2 is plotted against the degree of shrinkage; the thin gray line is as is in panel (c). With the optimized level of shrinkage identified via the *LW* estimator, there is a negligible loss in the model's predictive ability.



We performed ridge regression analyses separately for each imaging measure, for different "windows" or genomic regions of interest. In other words, we used a window-based scan of the genome, considering all nearby SNPs that are in high LD with a given SNP at the center of the window, as advocated by Malo et al. (2008). Windows were created by considering all SNPs that

passed a liberal, univariate GWAS *p*-value threshold of 0.10, along with their neighboring SNPs within a fixed distance, in single ridge regression models with the optimal shrinkage parameters. We tried a range of fixed window sizes (50 Kbp, 100 Kbp, 500 Kbp, and 1 Mbp) around the SNPs of interest. Here Kbp and Mbp denote thousands or millions of base pairs on the genome.

D. Multiple Comparisons

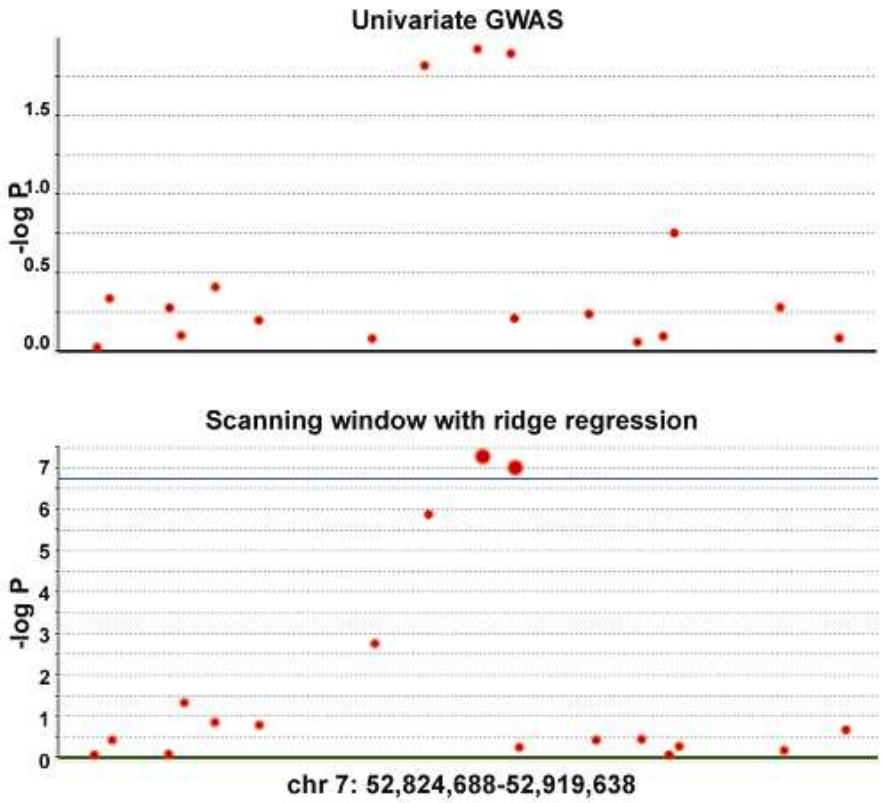
To correct for multiple comparisons, we divided the nominal *p*-value threshold of 0.05 by an estimate of the *effective* number of statistical tests across the genome. The effective number of tests (M_{eff}) was calculated for each chromosome using the simpleM program, as detailed in Stein et al. (2010b). To perform *post hoc*, exploratory tests on the top SNPs, we created voxelwise statistical maps using standard linear regression. To correct for multiple spatial comparisons, we used the standard False Discovery Rate method (FDR; Benjamini and Hochberg, 1995).

Results

As described in the methods, the corrected *p*-value threshold was set to be 0.05 divided by the effective number of tests (i.e., 264,889), which is 1.89×10^{-7}. None of the 437,607 SNPs in our study passed this threshold in association with hippocampal volume with standard, univariate GWAS. By using our genomic scanning of SNP windows using ridge regression, we were able to identify two intergenic SNPs (rs2912975 and rs4747490) that passed the stringent "genome-wide" significance threshold (i.e., correcting for all the statistical tests across the genome). As mentioned in the methods, several genomic window sizes were used, which yield different *p*-values for their corresponding SNPs. Below, we report the most significant *p*-values for the SNPs that passed the genome-wide *p*-value threshold.

The rs2912975 polymorphism on chromosome 7 was significantly associated with hippocampal volume (p-value = 4.98 x 10^{-8}), along with its neighboring SNP, with which it is almost perfectly correlated. The SNP is located close to a predicted, regulatory sequence of DNA. Univariate GWAS yielded a non-significant p-value of 1.19 x 10^{-2} for this SNP on chromosome 7. We plotted the ridge regression p-values from a scanning window with rs2912975 and its neighbors as an example of boosted genomic association power from ridge regression, along with the standard GWAS p-values for the same SNPs (**Figure 5.2**). In addition, another intergenic SNP, rs4747490, on chromosome 10, was significantly associated with hippocampal volume. In the temporal lobe volume analysis, our genomic scanning of SNP windows with ridge regression boosted the significance of the intergenic, likely regulatory rs2456930 SNP on chromosome 15 closer to the significance threshold.

Figure 5.2. The rs2912975 SNP has a *p*-value of 1.19 x 10^{-2} with standard, univariate GWAS, when correlated with hippocampal volume (*top panel*). In a ridge regression model, when considered along with SNPs in its vicinity, the same SNP has a genome-wide significant *p*-value of 4.98 x 10^{-8} (ridge regression *p*-values for a scanning window containing this SNP are displayed in the bottom panel; the top SNP is rs2912975).



Since the rs10845840 SNP in the *GRIN2B* gene is statistically significant in its association with temporal lobe volume with standard univariate GWAS (Stein et al., 2010), we expected its significance to be also detected with our genome-wide ridge regression scanning technique, possibly at a boosted level. This was not the case, however. Therefore, we looked more closely at the ridge regression windows centered at this SNP. We found that although rs10845840 was the top SNP at 50Kbp and 100Kbp scanning windows, another SNP in the *GRIN2B* gene (rs1805502) was the most significant SNP in the larger window sizes of 500Kbp and 1Mbp. In

fact, in the 500-Kbp window, rs1805502 had a more significant *p*-value (7.51×10^{-5}) than its univariate *p*-value of 9.08×10^{-5}. The rs1805502 SNP resides in the 3'-untranslated region (UTR) of the *GRIN2B* gene, almost 200Kbp away from rs10845840. The two SNPs are not in LD with each other, i.e. they are not statistically correlated in the population ($r^2 = 0.006$), and may therefore represent two independent contributions of the *GRIN2B* gene to the temporal lobe volume phenotype. To further explore this, we considered rs1805502 along with rs10845840 in a standard, multiple regression model (with no other SNPs). Both SNPs obtained *p*-values < 0.001, with minor alleles having effects in opposite directions. All other SNPs around rs10845840, even in our largest window size of 1Mbp, had *p*-values > 0.001 when they were paired with rs10845840 in standard, multiple regression models.

In *post hoc*, exploratory tests, we evaluated more closely the effects of the genetic variants that ridge regression found to be strongly associated with our MRI-based summary measures. We performed voxel-by-voxel association studies using TBM. SNPs were coded in an additive fashion (0, 1 or 2 for the number of minor alleles). Volumetric differences at each voxel were correlated with the SNPs separately using standard regression, after adjusting for age and sex. We found that rs2456930 showed significant effects (FDR critical *p*-value of 0.023) in the temporal lobe voxels, using both temporal lobes as the search regions of interest (**Figure 5.4**), and at the whole-brain level, with an FDR critical *p*-value of 3.71×10^{-3} (i.e., the highest *P* value threshold that controls the FDR at the conventional 5% rate). Furthermore, the rs2912992 SNP had a significant association (with an FDR critical *p*-value of 0.0099) selectively within the left hippocampal voxels (**Figure 5.3**), although it did not pass FDR for both hippocampi or for the

right hippocampus. Replication studies and meta-analyses in even larger samples are underway, to confirm this.

Figure 5.3. TBM reveals the profile of statistically significant effects of rs2912975 on the left hippocampus. FDR was used to correct for multiple comparisons in the left hippocampal voxels, using a binary mask. The image is in radiological convention (the right side of the image shows the left side of the subject's brain).

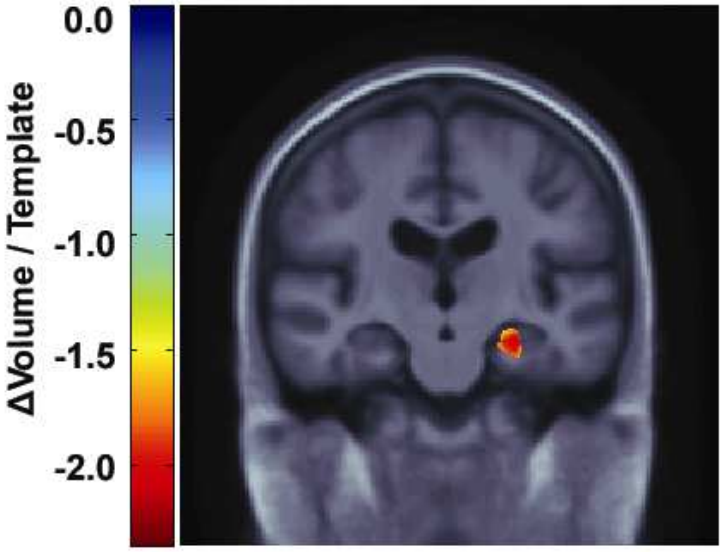
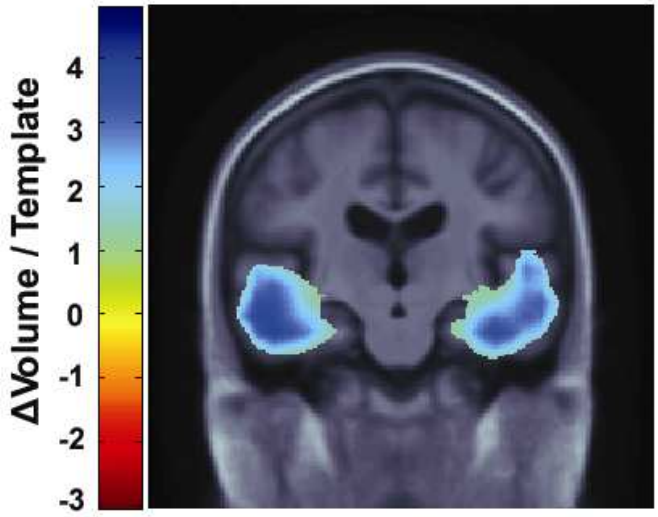


Figure 5.4. TBM reveals the profile of statistically significant effects of the rs2456930 polymorphism on the temporal lobes. FDR was used to correct for multiple comparisons in the temporal lobe voxels, using a binary mask for the temporal lobes. The image is in radiological convention (the right side of the image shows the left side of the subject's brain).



Conclusion

We applied a novel, genome-wide, ridge regression approach to study the association of multiple SNPs with imaging phenotypes. We identified three SNPs with significant or near genome-wide significant effects on MRI-derived hippocampal volume and temporal lobe volume measures. In some but not all cases, associations were boosted in power relative to standard, univariate genome-wide association analyses. In addition, two of the three SNPs we identified had significant voxelwise effects in *post hoc* analyses and are located near regulatory DNA sequences, making them potentially important genetic variants for influencing brain structure in large populations.

References

Thompson PM, Martin NG, Wright MJ. Imaging genomics. Curr Opin Neurol. 2010; 23(4):368-373.

Stein JL, Hua X, Morra JH, Lee S, Hibar DP, Ho AJ, Leow AD, Toga AW, Sul JH, Kang HM, Eskin E, Saykin AJ, Shen L, Foroud T, Pankratz N, Huentelman MJ, Craig DW, Gerber JD, Allen AN, Corneveaux JJ, Stephan DA, Webster J, DeChairo BM, Potkin SG, Jack CR Jr, Weiner MW, Thompson PM; Alzheimer's Disease Neuroimaging Initiative. Genome-wide analysis reveals novel genes influencing temporal lobe structure with relevance to neurodegeneration in Alzheimer's disease. NeuroImage. 2010; 51(2):542-554.

Potkin SG, Guffanti G, Lakatos A, Turner JA, Kruggel F, Fallon JH, Saykin AJ, Orro A, Lupoli S, Salvi E, Weiner M, Macciardi F; Alzheimer's Disease Neuroimaging Initiative. Hippocampal atrophy as a quantitative trait in a genome-wide association study identifying novel susceptibility genes for Alzheimer's disease. PLoS One. 2009; 4(8):e6501.

The ENIGMA Consortium. Genome-Wide Association Meta-Analysis of Hippocampal Volume: Results from the ENIGMA Consortium. Organization for Human Brain Mapping meeting. Quebec City, Canada. 2011.

Szymczak S, Biernacka JM, Cordell HJ, González-Recio O, König IR, Zhang H, Sun YV. Machine learning in genome-wide association studies. Genet Epidemiol. 2009;33(Suppl 1):S51-7.

Sun YV, Shedden KA, Zhu J, Choi NH, Kardia SL. Identification of correlated genetic variants jointly associated with rheumatoid arthritis using ridge regression. BMC Proc. 2009; 3(Suppl):7-S67.

Kim Y, Wojciechowski R, Sung H, Mathias RA, Wang L, Klein AP, Lenroot RK, Malley J, Bailey-Wilson JE. Evaluation of random forests performance for genome-wide association studies in the presence of interaction effects. BMC Proc. 2009; 3(Suppl):7-S64.

Hoerl AE. Optimum solution of many variable equations. Chem Eng Prog. 1959; 55:67-78.

Hoerl AE, Application of ridge analysis to regression problems. Chem Eng Prog. 1962; 58:54-59.

Hoerl AE, Kennard RW. Ridge regression: Biased estimation for non-orthogonal problems. Technometrics 1970; 12:55-67.

Malo N, Libiger O, Schork NJ. Accommodating linkage disequilibrium in genetic-association analyses via ridge regression. Am J Hum Genet. 2008; 82(2):375-385.

Morra JH, Tu Z, Apostolova LG, Green AE, Avedissian C, Madsen SK, Parikshak N, Hua X, Toga AW, Jack CR Jr, Weiner MW, Thompson PM; Alzheimer's Disease Neuroimaging Initiative. Validation of a fully automated 3D hippocampal segmentation method using subjects with Alzheimer's disease mild cognitive impairment, and elderly controls. NeuroImage. 2008; 43(1):59-68.

Hua X, Leow AD, Lee S, Klunder AD, Toga AW, Lepore N, Chou YY, Brun C, Chiang MC, Barysheva M, Jack CR Jr, Bernstein MA, Britson PJ, Ward CP, Whitwell JL, Borowski B, Fleisher AS, Fox NC, Boyes RG, Barnes J, Harvey D, Kornak J, Schuff N, Boreta L, Alexander GE, Weiner MW, Thompson PM, Alzheimer's Disease Neuroimaging Initiative. 3D characterization of brain atrophy in Alzheimer's disease and mild cognitive impairment using tensor-based morphometry. NeuroImage. 2008; 41(1):19-34.

Saykin AJ, Shen L, Foroud TM, Potkin SG, Swaminathan S, Kim S, Risacher SL, Nho K, Huentelman MJ, Craig DW, Thompson PM, Stein JL, Moore JH, Farrer LA, Green RC, Bertram L, Jack CR Jr, Weiner MW; Alzheimer's Disease Neuroimaging Initiative. Alzheimer's Disease Neuroimaging Initiative biomarkers as quantitative phenotypes: Genetics core aims, progress, and plans. Alzheimers Dement. 2010; 6(3):265-273.

Stein JL, Hua X, Lee S, Ho AJ, Leow AD, Toga AW, Saykin AJ, Shen L, Foroud T, Pankratz N, Huentelman MJ, Craig DW, Gerber JD, Allen AN, Corneveaux JJ, Dechairo BM, Potkin SG, Weiner MW, Thompson P; Alzheimer's Disease Neuroimaging Initiative. Voxelwise genome-wide association study (vGWAS). NeuroImage. 2010; 53(3):1160-1174.

Tikhonov AN, Arsenin VY. Solutions of ill-posed problems. Winston, Washington. 1977.

Donoho DL. Compressed sensing. IEEE Trans Inf Theory. 2006; 52:1289-1306.

Goldstein T, Osher S. The split Bregman method for L1 regularized problems. UCLA CAM report. 2008; 8-29.

Chatterjee S, Price B. Regression analysis by example. Wiley, New York. 1977.

Halawa AM, El Bassiouni MY. Tests of regression coefficients under ridge regression models. J Statist Comput Stimul. 2000; 65:341-356.

Hoerl AE, Kannard RW, Baldwin KF. Ridge regression: Some simulations. Comm Statist Theory Methods. 1975; 4:105-123.

Lawless JF, Wang P. A simulation study of ridge and other regression estimators. Commun Stat A- Theor. 1976; 5(4):307-323.

Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. J R Statist Soc B. 1995; 57(1):289-300.

Chapter 6. Prediction of temporal lobe volume on MRI from genotypes with elastic net

Introduction

Many early studies in imaging genetics explored univariate associations between genotypes and imaging measures, assuming each gene acted independently. One disadvantage of such studies is their limited statistical power to detect gene effects on the brain. Meta-analyses such as the Enhancing Neuro Imaging Genetics through Meta-Analysis (ENIGMA) project (The ENIGMA Consortium, 2011) have boosted statistical power, by analyzing MRI and genome-wide genotype data from over 20,000 subjects, gaining power from very large sample sizes. Multivariate approaches, which simultaneously consider entire sets of genotypes, sets of voxels in an image, or both, have also become more popular (Hibar et al. 2011a), as they also handle potential problems in high-dimensional data, such as highly correlated predictors, where almost all have no detectable effects.

In Hibar et al. (2011a), we reviewed several recent multivariate, imaging genetics studies that applied principal component regression (Hibar et al., 2011b), sparse reduced rank regression (Vounou et al., 2010), or independent components analysis (Liu et al., 2009) to discover genetic influences on the brain that would have been missed by using only univariate techniques. Regularized, sparse regression methods, in particular, use penalty terms to tackle the problems of high dimensionality (e.g., having more predictors than samples), multiple highly correlated measures, and multiple comparisons across an image, the genome, or both. The "elastic net" combines L^1- and L^2-norm regularization and benefits from the advantages of both methods, to handle high-dimensional, highly correlated data. The algorithm takes advantage of the sparsity

properties of L^1 (Least Absolute Shrinkage and Selection Operator, or LASSO), along with the stability of L^2 (ridge) regression (Zou and Hastie, 2005). Here, we introduce an elastic net approach to predict an imaging measure from top genotypes. We aim to incorporate top genetic variants (i.e., single nucleotide polymorphisms or SNPs), screened based on univariate genome-wide search (as in a genome-wide association analysis or GWAS), into an elastic net model, to predict temporal lobe volume on MRI. Recently, the elastic net has been applied to genomics (Cho et al., 2009; Cho et al., 2010) for jointly considering genetic polymorphisms as well as imaging (Bunea et al., 2011), to integrate large numbers of imaging and clinical predictors. More recently, the algorithm has also been used to detect multi-SNP associations with hippocampal surface morphometry (Wan et al., 2011), and to integrate imaging and proteomic data in Alzheimer's disease (Shen et al., 2011).

We hypothesize that this doubly regularized, multivariate regression method would allow us to make significant predictions of MRI-derived temporal lobe volume from genotypes. This predictive approach, we propose, may have implications for early, personalized risk assessment of brain disorders such as Alzheimer's disease, where the temporal lobes undergo significant atrophy.

Methods

A. MRI Measures

ADNI subjects were scanned with a standard MRI protocol optimized for reproducibility and consistency across 58 sites in North America. Temporal lobe volumes were derived from an anatomically defined region-of-interest (ROI) on three-dimensional maps of relative volumes

generated with tensor-based morphometry (TBM), a well-established method to map volumetric differences in the brain (Hua et al., 2008). Temporal lobe volume is particularly interesting, as this structure is prone to atrophy in Alzheimer's disease (AD). There is interest in discovering genes that may promote or resist the atrophy, or contribute to normal variations in its volume. A total of 740 subjects with both imaging and genotype data were included (173 with AD, 361 with mild cognitive impairment or MCI, and 206 cognitively healthy controls; 438 men and 302 women; mean ± SD age: 75.55 ± 6.79 years).

B. Genotypes

Genotyping procedures for ADNI are described in Saykin et al. (2010). SNPs with minor allele frequencies less than 0.01 and Hardy-Weinberg equilibrium *p*-values less strict than 5.7×10^{-7} were excluded. Genotypes were imputed to infer missing information.

C. Elastic net method

The elastic net (Zou and Hastie, 2005) is a form of penalized regression, where both L^1 and L^2 regularizations are introduced into the standard multiple linear regression model, as formulated below for *n* subjects and *p* predictors:

$$\beta^* = \arg\min_{\beta} \|y - X\beta\|^2 + \lambda_1 \|\beta\|_1 + \lambda_2 \|\beta\|^2 \qquad (1)$$

Here, y represents the vector whose n components are the imaging measure for each subject, after adjusting for sex and age (residuals of regression). X is the n x p matrix of genotypes for top genetic variants across the genome. β^* represents the vector of fitted regression coefficients for each SNP's effect on the imaging measure. λ_1 is a positive weighting parameter on the L^1 penalty, which promotes sparsity in the resulting set of fitted regression coefficients, as many

coefficients are likely to be exactly zero. λ_2 is a positive weighting parameter on the L^2 penalty, which promotes stability in the regularization path and precludes a limit on how many variables are selected (in strict LASSO, at most *n* variables can be selected in an *n* by *p* case).

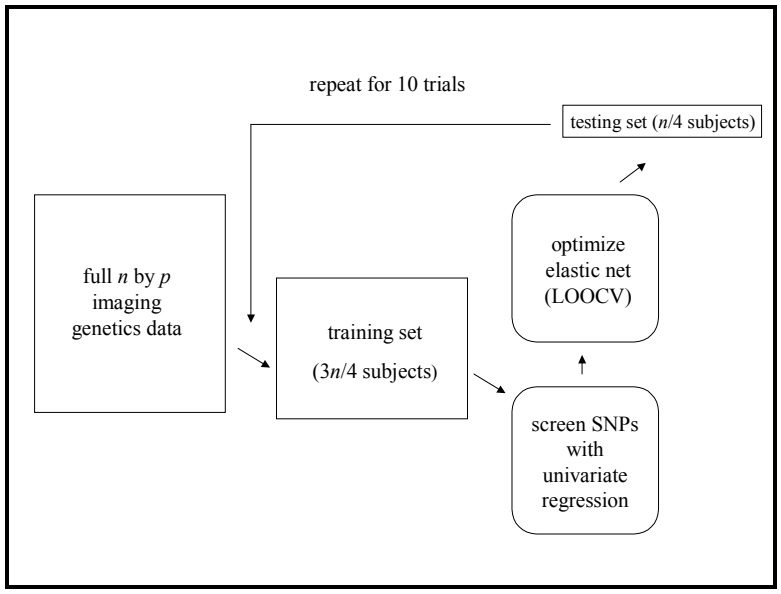
In ten separate experiments (**Figure 6.1**), we randomly split the data into training sets with 3*n*/4 and testing sets with *n*/4 subjects. Standard univariate associations were performed for all ~500,000 genotyped variants with the imaging measure, using the training set only, and top 4,000 SNPs were then fed into the elastic net algorithm. This is a common pre-screening step that has been used in similar contexts (Cho et al., 2010). Leave-one-out cross-validation was performed within the training sets to determine the optimal penalty parameters with the mean squared error criterion. Both λ_1 and α are optimized with a grid search, where $a = \lambda_2 / (\lambda_1 + \lambda_2)$, such that the penalty term of (1), *P*, is restated as below:

$$P = \alpha\|\beta\|^2 + (1-\alpha)\|\beta\|_1 \qquad (2)$$

Mean squared error is commonly minimized for parameter tuning using cross-validation, similarly to previous studies in this context (Shen et al., 2011; Wan et al., 2011). To avoid bias, cross-validation for selecting hyperparameters is done separately from evaluation of the model. Models trained to have optimal penalty parameters were tested on the test sets to obtain mean squared errors for predicting the imaging measure from genotypes. For our analyses, we used the 'glmnet' package (Friedman et al., 2010) implemented in R (http://cran.r-project.org). This optimizes model fitting parameters via an efficient, coordinate descent algorithm.

A similar procedure was repeated 100,000 times. To reduce computational time, unlike the actual experiments, only the optimal penalty parameters were used and a fixed set of top 4,000 SNPs from a univariate genome-wide search were incorporated into the models. Imaging measures were randomly assigned to all subjects, after which the data was randomly split into training and testing sets as above. Mean squared errors for prediction of test set temporal lobe volumes were then obtained for each permutation.

Figure 6.1. Validation framework. Different loops of cross-validation are necessary to prevent over-fitting of a predictive model. We pre-screen the single nucleotide polymorphisms (SNPs) for dimension reduction, and elastic net parameter optimization, is only performed within the training data. The mean squared errors of predictions in 10 separate trials on independent test sets are averaged. LOOCV = Leave-one-out cross-validation.



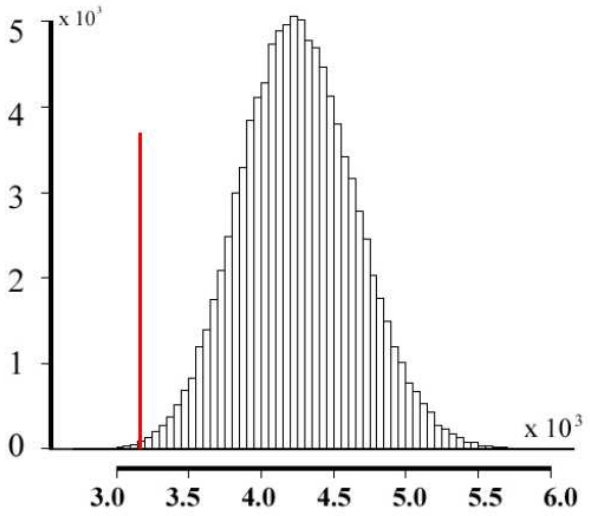
Standard multiple regression cannot be used in our scenario, as the multivariate analysis for all top SNPs would fail (i.e., the model fitting equation would be ill-conditioned), as there are many more variants than subjects ($p >> n$ problem).

To perform *post-hoc*, exploratory tests on our top SNPs, we created voxelwise statistical maps to reveal the spatial profile of associations with regional brain volumes. We fitted linear associations at each voxel, adjusted for covariates (sex and age). To correct for multiple spatial comparisons, we used a regional False Discovery Rate (FDR) method, which is now fairly standard in neuroimaging (Langers et al., 2007).

Results

We averaged the mean squared errors of the optimized predictive models on test sets. An average mean squared error of 3,147 was obtained with the elastic net predictor in independent sets of test subjects. The average mean squared error in the 100,000 permutations was 4,257 with a standard deviation of 397. Compared to the distribution of the errors across the permutations (**Figure 6.2**), the *p*-value is found to be close to 0.001.

Figure 6.2. Distribution of mean squared errors for the 10^5 simulations conducted with the optimal elastic net parameters. Errors are approximately normally distributed (mean, 4,257; SD: 397). 131 permutations had errors smaller than our predictive model's error (*red line*), yielding an empirical *p*-value ~ 0.001.



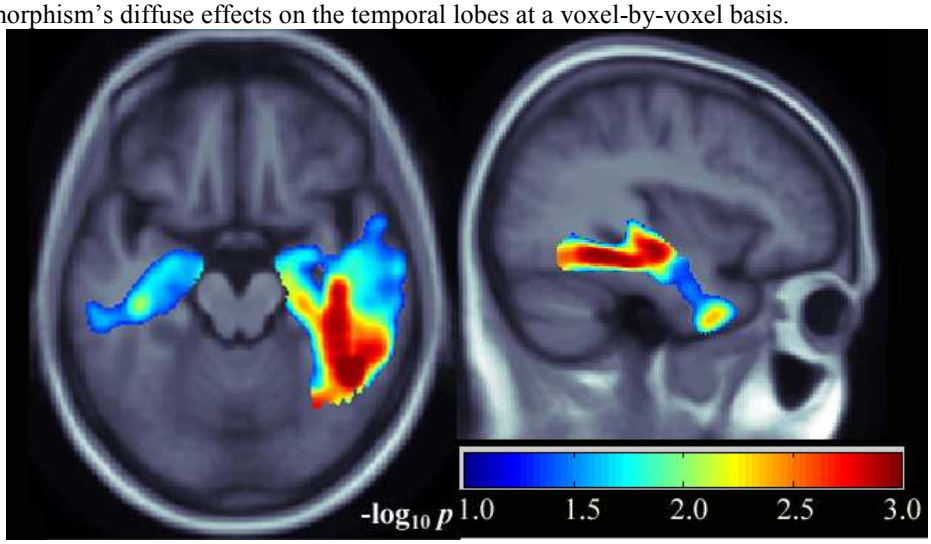
To investigate which genetic variants contributed most to the predictions, we examined the average absolute values of coefficients for each fitted predictor. Out of the 4,000 variants incorporated into the elastic net models in each of the ten trials, 105 were screened for all trials. We investigated the coefficients obtained by these SNPs. The top ten are shown in **Table 6.1**. To ensure that the findings were robust, we also counted the number of times the variants received nonzero coefficients across the ten runs (**Table 6.1**). With permutations, each SNP obtained a nonzero coefficient only about 2.0 ± 0.5 SD times, on average.

Table 6.1. List of single nucleotide polymorphisms (SNPs) with the highest contribution to the elastic net models predicting temporal lobe volume on MRI. These ten SNPs had the largest elastic net coefficients (absolute values), and their selection was robust, as they obtained nonzero coefficients at least 8 out of the 10 total trials. Corresponding gene names and chromosome numbers are displayed for the variants.

SNP	**Gene**	**Chr**	$\|\beta\|_{average}$	**$\|\beta\|$>0 count**
rs2456930	-	15	2.32	10
rs10518480	-	4	1.96	10
rs17476752	-	5	1.78	9
rs9933137	*RBFOX1*	16	1.75	8
rs10845840	*GRIN2B*	12	1.64	9
rs997972	-	20	1.50	9
rs1929933	*GLDC*	9	1.44	9
rs1564348	*SLC22A1*	6	1.41	9
rs309800	-	4	1.37	10
rs11204135		8	1.33	10

We noted that rs10845840 in the *GRIN2B* gene and the intergenic rs2456930, which were the top findings with a univariate genome-wide search (Stein et al., 2010), also appeared in our top list, which is a re-assuring validation. Interestingly, rs9933137 in the *RBFOX1* gene also obtained a very high mean $|\beta|$ and outperformed the top univariate SNP in *GRIN2B*. To explore the profile of effects of the *RBFOX1* SNP on temporal lobes in more detail, we performed an exploratory, *post-hoc* voxelwise test, shown in **Figure 6.3**.

Figure 6.3. The *post-hoc* voxelwise effects of the *RBFOX1* rs9933137 polymorphism are shown on TBM-derived maps of the temporal lobes, using linear regression. Volumetric change at each voxel is linearly regressed against the genetic variant, along with covariates such as sex and age. *P*-values for the associations are corrected for multiple spatial comparisons using regional false discovery rate (FDR). Warmer colors represent more significant effects. Images are in radiological convention. Results survived multiple comparisons correction across both lobes, but the left temporal lobe showed stronger effects (also seen in the left sagittal slice). Although this does not add new information to the multivariate, prediction study, it confirms that the highly predictive polymorphism's diffuse effects on the temporal lobes at a voxel-by-voxel basis.



Conclusion

We proposed a multivariate model to predict an imaging measure from genotypes, using L^1-L^2 regularized regression, also known as the elastic net. We split 740 ADNI subjects into training and test sets in ten separate trials. We optimized elastic net parameters in the training set using

leave-one-out cross-validation, and predictions were made on the independent test sets. This is a rigorous predictive framework, as it avoids the overfitting that can arise if training data are used for testing. We also compared the performance of our predictor with that of 10^5 permutations, where MRI measures were randomly assigned to the subjects. Our predictions were significantly better than those made by random models. Although the main goal of our study was prediction rather than discovery, we also looked for the variants that most strongly contributed to the predictions. Using average elastic net coefficients as a metric, we found a single nucleotide polymorphism in the *RBFOX1* gene to be most contributory to the predictive models, which also showed significant 3D effects on the temporal lobes. This gene, also known as *A2BP1*, has been previously characterized as an autism risk gene (Martin et al., 2007), and regulates neuronal excitation in the brain (Gehman et al., 2011). Interestingly, it has also been discovered in another sparse regression imaging genetics study as a highly significant gene (Vounou et al., 2011). Future studies are needed to compare the performance of this predictor with other multivariate techniques. Pre-screening of genetic variants, which was done as a way of reducing dimensionality similarly to previous studies (Cho et al., 2010), may be a limitation, as it might lead to missing potential effects from contributory genes. Furthermore, applying multi-voxel methods (Liu et al., 2009; Vounou et al., 2010; Vounou et al., 2011) and incorporating biological pathway information may yield more statistically powerful predictions.

References

The ENIGMA Consortium. Genome-Wide Association Meta-Analysis of Hippocampal Volume: Results from the ENIGMA Consortium. Organization for Human Brain Mapping meeting. Quebec City, Canada. 2011.

Hibar DP, Kohannim O, Stein JL, Chiang MC, Thompson PM. Multilocus genetic analysis of brain images. Front Genet. 2011; 2:73.

Hibar DP, Stein JL, Kohannim O, Jahanshad N, Saykin AJ, Shen L, Kim S, Pankratz N, Foroud T, Huentelman MJ, Potkin SG, Jack CR Jr, Weiner MW, Toga AW, Thompson PM; Alzheimer's Disease Neuroimaging Initiative. Voxelwise gene-wide association study (vGeneWAS): multivariate gene-based association testing in 731 elderly subjects. NeuroImage. 2011; 56(4):1875-1891.

Vounou M, Nichols TE, Montana G; Alzheimer's Disease Neuroimaging Initiative. Discovering genetic associations with high-dimensional neuroimaging phenotypes: A sparse reduced-rank regression approach. NeuroImage. 2010; 53(3):1147-1159.

Liu J, Pearlson G, Windemuth A, Ruano G, Perrone-Bizzozero NI, Calhoun V. Combining fMRI and SNP data to investigate connections between brain function and genetics using parallel ICA. Hum Brain Mapp. 2009; 30(1):241-255.

Zou H, Hastie T. Regularization and variable selection via the elastic net. J Royal Stat Soc B. 2005; 67(2):301-320.

Cho S, Kim K, Kim YJ, Lee JK, Cho YS, Lee JY, Han BG, Kim H, Ott J, Park T. Joint identification of multiple genetic variants via elastic-net variable selection in a genome-wide association analysis. Ann Hum Genet. 2010; 74(5):416-428.

Cho S, Kim H, Oh S, Kim K, Park T. Elastic-net regularization approaches for genome-wide association studies of rheumatoid arthritis. BMC Proc. 2009; 3(Suppl):7-S25.

Bunea F, She Y, Ombao H, Gongvatana A, Devlin K, Cohen R. Penalized least squares regression methods and applications to neuroimaging. NeuroImage. 2011; 55(4):1519-1527.

Wan J, Kim S, Inlow M, Nho K, Swaminathan S, Risacheri SL, Fang S, Weiner MW, Beg MF, Wang L, Saykin AJ, Shen L; Alzheimer's Disease Neuroimaging Initiative. Hippocampal surface mapping of genetic risk factors in AD via sparse learning models. Med Image Comput Comput Assist Interv. 2011; 14(Pt 2):376-383.

Shen L, Kim S, Qi Y, Inlow M, Swaminathan S, Nho K, Wan J, Risacher SL, Shaw LM, Trojanowski JQ, Weiner MW, Saykin AJ and the Alzheimer's Disease Neuroimaging Initiative. Identifying Neuroimaging and Proteomic Biomarkers for MCI and AD via the Elastic Net. Lect Notes Comput Sci. 2011; 7012:27-34.

Hua X, Leow AD, Lee S, Klunder AD, Toga AW, Lepore N, Chou YY, Brun C, Chiang MC, Barysheva M, Jack CR Jr, Bernstein MA, Britson PJ, Ward CP, Whitwell JL, Borowski B, Fleisher AS, Fox NC, Boyes RG, Barnes J, Harvey D, Kornak J, Schuff N, Boreta L, Alexander GE, Weiner MW, Thompson PM, Alzheimer's Disease Neuroimaging Initiative. 3D characterization of brain atrophy in Alzheimer's disease and mild cognitive impairment using tensor-based morphometry. NeuroImage. 2008; 41(1):19-34.

Saykin AJ, Shen L, Foroud TM, Potkin SG, Swaminathan S, Kim S, Risacher SL, Nho K, Huentelman MJ, Craig DW, Thompson PM, Stein JL, Moore JH, Farrer LA, Green RC, Bertram L, Jack CR Jr, Weiner MW; Alzheimer's Disease Neuroimaging Initiative. Alzheimer's Disease Neuroimaging Initiative biomarkers as quantitative phenotypes: Genetics core aims, progress, and plans. Alzheimers Dement. 2010; 6(3):265-273.

Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw. 2010; 33(1):1-22.

Langers DR, Jansen JF, Backes WH. Enhanced signal detection in neuroimaging by means of regional control of the global false discovery rate. NeuroImage. 2007; 38(1):43-56.

Stein JL, Hua X, Morra JH, Lee S, Hibar DP, Ho AJ, Leow AD, Toga AW, Sul JH, Kang HM, Eskin E, Saykin AJ, Shen L, Foroud T, Pankratz N, Huentelman MJ, Craig DW, Gerber JD, Allen AN, Corneveaux JJ, Stephan DA, Webster J, DeChairo BM, Potkin SG, Jack CR Jr,

Weiner MW, Thompson PM; Alzheimer's Disease Neuroimaging Initiative. Genome-wide analysis reveals novel genes influencing temporal lobe structure with relevance to neurodegeneration in Alzheimer's disease. NeuroImage. 2010; 51(2):542-554.

Martin CL, Duvall JA, Ilkin Y, Simon JS, Arreaza MG, Wilkes K, Alvarez-Retuerto A, Whichello A, Powell CM, Rao K, Cook E, Geschwind DH. Cytogenetic and molecular characterization of A2BP1/FOX1 as a candidate gene for autism. Am J Med Genet B Neuropsychiatr Genet. 2007; 144B(7):869-76.

Gehman LT, Stoilov P, Maguire J, Damianov A, Lin CH, Shiue L, Ares M Jr, Mody I, Black DL. The splicing regulator Rbfox1 (A2BP1) controls neuronal excitation in the mammalian brain. Nat Genet. 2011; 43(7):706-711.

Vounou M, Janousova E, Wolz R, Stein JL, Thompson PM, Rueckert D, Montana G; Alzheimer's Disease Neuroimaging Initiative. Sparse reduced-rank regression detects genetic associations with voxel-wise longitudinal phenotypes in Alzheimer's disease. NeuroImage. 2012; 60(1):700-716.

Chapter 7. Multi-gene prediction of white matter structure

Introduction

Diffusion tensor imaging (DTI) is widely acknowledged as a useful tool for studying the white matter microstructure of the living brain. By mapping the diffusion of water through the brain's fibers, DTI can recover major fiber pathways in the brain, and patterns of anatomical connectivity, with broad applications in psychiatry, neurology and brain mapping (Thomason and Thompson, 2011). DTI-based white matter abnormalities are widely reported in developmental and degenerative brain diseases including Alzheimer's disease and mild cognitive impairment (Fellgiebel *et al*, 2004; Naggara *et al*, 2006; Oishi *et al*, 2011), schizophrenia (White *et al*, 2008; Ellison-Wright and Bullmore, 2009; Patel *et al*, 2011), bipolar disorder (Sussman *et al*, 2009; Heng *et al*, 2010), attention-deficit/hyperactivity disorder (ADHD; Konrad *et al*, 2010), and autism (Alexander *et al*, 2007; Ke *et al*, 2009). These studies show the utility of DTI in neuropsychiatric research. In several studies, treatment of neuropsychiatric patients has also been associated with changes in DTI measures (Versace *et al*, 2008; Yoo *et al*, 2007). This also shows the promise of DTI for understanding therapeutic effects.

Measures of white matter integrity derived from DTI, such as fractional anisotropy (FA), are highly heritable (Lee *et al*, 2008; Chiang *et al*, 2009; Kochunov *et al*, 2010; Lee *et al*, 2010; Patel *et al*, 2010; Chiang *et al*, 2011b). As such, they may be useful as intermediate measures or "endophenotypes" (Meyer-Lindenberg and Weinberger, 2006; de Geus *et al*, 2008; Hall and Smoller, 2010; Marenco and Radulescu, 2010) for assessing genetic influences on the brain. Several commonly-carried genetic variants have already been identified that exert small effects

on the brain's white matter as detected by DTI. These include highly prevalent polymorphisms in genes coding for brain-derived neurotrophic factor (*BDNF*; Chiang *et al*, 2011a), clusterin (*CLU*; Braskie *et al*, 2011b), the neuregulin 1 receptor (*ErbB4*; Konrad *et al*, 2009), neurotrophic tyrosine kinase receptor, type 1 (*NTRK1*; Braskie *et al*, 2011a), catechol-*O*-methyl transferase (*COMT*; Thomason *et al*, 2010), and the hemochromatosis gene, *HFE* (Jahanshad *et al*, 2012a). We therefore considered these genes as candidates in this study.

The molecular and cellular effects of these genes and their protein products have been extensively investigated. *COMT* is a well-studied gene and codes for one of a group of enzymes that degrade catecholamines. Catecholamine levels are altered in many neuropsychiatric disorders, thereby making this molecule an ideal target for medications. Several of the genes above are also well-known for their role in brain development. *BDNF*'s protein product is a neural growth factor or *neurotrophin*, vital for the healthy development and maintenance of the nervous system (Binder and Scharfman, 2004). Similarly, *NTRK1* codes for TrkA, which belongs to a tyrosine kinase receptor family, to which neurotrophin growth factors bind. Neurotrophins and their receptors, not surprisingly, are also important in neuropsychiatric disease and may offer new therapeutic targets in the form of small-molecule antagonists or mimickers (Allen and Dawbarn, 2006). *ErbB4* encodes another tyrosine kinase receptor, which by binding to its ligand, neuregulin-1 (coded by *NRG1*), participates in neural modulation and development and is thought to contribute to the pathophysiology of schizophrenia (Hahn *et al*, 2006). Lastly, *HFE* and *CLU* contain polymorphisms that increase the risk for neurodegenerative disease. Their protein products regulate iron metabolism – important in brain aging (Bartzokis *et al*, 2011) – and beta-amyloid metabolism (DeMattos *et al*, 2002), respectively.

White matter structure is certainly influenced by non-genetic factors such as age (Chiang *et al*, 2011b), and sex differences (which are partly genetic and nongenetic), but we expect a moderate and significant proportion of an individual's white matter integrity to be predictable from their genetic profiles. This is corroborated by DTI findings of high heritability for white matter microstructure. As mentioned above, individual effects of single genetic variants on white matter structure have been explored, but a *multilocus* approach has not yet been taken. The utility of a multilocus candidate gene model in predicting an imaging-derived outcome was recently explored in the context of structural MRI (Biffi *et al*, 2010) and functional MRI (Nikolova *et al*, 2011), but its applications in diffusion tensor imaging and detailed three-dimensional maps of brain structure appear novel. In this paper, we incorporate a subject's genetic signature, at key loci, into a multilocus model. We hypothesize that this would help predict brain integrity, as measured by DTI-derived FA, more powerfully than a single-locus genetic test. We focus on the corpus callosum, as it is the largest white matter structure in the brain, easy to examine at the brain's midline, highly heritable (Chiang *et al*, 2009; Brouwer *et al*, 2010; Kochunov *et al*, 2010), and well-studied in neurology and psychiatry as the primary commissure connecting the two brain hemispheres (Foong *et al*, 2000; Alexander *et al*, 2007). We chose FA as the DTI measure of white matter structure, as it has been shown to have higher heritability than other DTI parameters, such as radial and axial diffusivity (Kochunov *et al*, 2010).

Methods

A. Participants

A total of 395 subjects (23.7 ± 2.2 years of age; 143 men and 252 women; 47 siblings, 141 monozygotic twins (49 pairs and 43 singles) and 207 dizygotic twins (1 triplet, 70 pairs and 64 singles) from the Brisbane young adult twins and siblings study (de Zubicaray *et al*, 2008) were included in our study, for whom both 105-gradient DTI scans and genome-wide genotype information were available. All twins in this study are Australians of European descent. Previously, principal component analysis was conducted in this cohort for population stratification analysis and correction (Medland *et al*, 2009). Subjects who were more than six standard deviations from either of the top two average reference principal component scores – derived from non-Australian European populations – were identified as ancestry outliers and excluded from analysis. The first two principal components refer to differences between Africans and non-Africans and to differences between East Asians and others, respectively. Due to migration patterns and the fact that this sample was originally recruited to study mole patterns, exclusions are usually due to Asian or Polynesian ancestry. All subjects were screened to exclude cases of pathology known to affect brain structure. Additionally, no subjects had a first-degree relative with a psychiatric disorder or reported a history of significant head injury, a neurological or psychiatric illness, substance abuse or dependence.

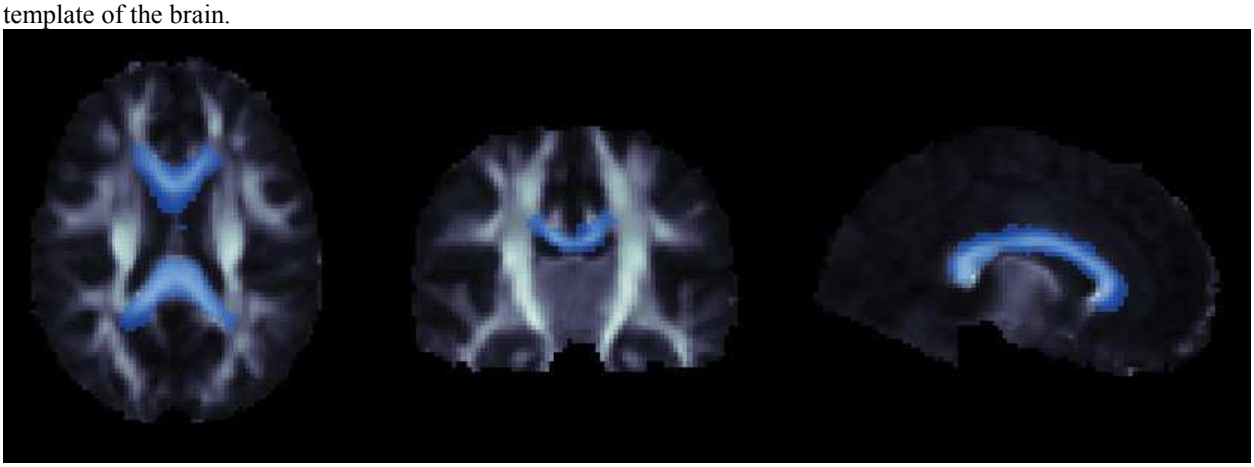
B. Diffusion Tensor Imaging

Whole-brain diffusion tensor MRI scans were collected with a 4-Tesla Bruker Medspec MRI scanner. Images were acquired using single-shot echo planar imaging with a twice-refocused spin echo sequence to reduce eddy-current induced distortions. Acquisition parameters were optimized to yield the best signal-to-noise ratio for estimation of diffusion tensors (Jones *et al*, 1999). Imaging parameters were: 23 cm field-of-view, TR/TE 6090/91.7 ms, with a 128 × 128

acquisition matrix. 105 images were acquired for each subject: 11 with no diffusion sensitization and 94 diffusion-weighted images with gradient directions evenly distributed on the hemisphere. Standard protocols for skull-stripping and eddy current distortion correction are performed using FSL (http://www.fmrib.ox.ac.uk/fsl) and we adjusted for echo planar imaging distortions as detailed in prior studies (Leow *et al*, 2005; Jahanshad *et al*, 2010). FSL was also used to calculate tensors and scalar maps of fractional anisotropy (FA) from the corrected images. The LONI pipeline (http://pipeline.loni.ucla.edu) was used to parallelize the preprocessing steps.

A mean deformation template (MDT) was created for the DTI scans, to which subjects' FA maps (obtained from DWI elastically aligned to their high resolution T1-weighted anatomical scan) were registered as in Jahanshad *et al* (2010), using a 3D elastic warping technique with a mutual information cost function (Leow *et al*, 2005). The MDT and the registered FA maps were then thresholded at 0.25, as FA measures below this threshold may reflect contributions from non-white matter in healthy-appearing white matter. After registering the FA maps across subjects, all FA images were smoothed with a Gaussian filter with a 7-mm isotropic full-width at half max (FWHM). The structure of the corpus callosum was identified automatically by using the Johns Hopkins University (JHU) white matter atlas (ICBM DTI 81; Mori *et al*, 2008), which tracks its 3D extent, extending laterally from the midline (**Figure 7.1**). The atlas FA image was linearly and then elastically registered to our study specific FA-MDT; the transformation and deformation map was then applied to the JHU set of labels using nearest-neighbor interpolation to avoid intermixing of labels. The full label (composed of three regions: *splenium*, body, and *genu*) of the corpus callosum was then accurately extracted. This avoided subjectivity and rater dependency in defining the limits of the corpus callosum.

Figure 7.1. The three-dimensional structure of the corpus callosum, as defined by the Johns Hopkins University (JHU) white matter atlas is displayed in axial, coronal and sagittal views in *blue*, overlaid on a mean deformation template of the brain.



C. Genotyping and Selection of Candidate SNPs

We considered six candidate SNPs listed in **Table 7.1** due to the recent imaging genetics discoveries outlined in the introduction. These particular genetic variants are located in six different genes. All have been linked to structural differences detectable with DTI. Several of these polymorphisms (i.e., rs6265 in *BDNF*, rs6336 in *NTRK1*, rs4680 in *COMT* and rs1799945 in *HFE*) are exonic variants and lead to amino acid changes in the protein products of these genes (val→met, his→tyr, val→met and his→asp, respectively). These have been well-studied in the neuropsychiatric literature (Egan *et al*, 2003; Zecca *et al*, 2004; Tunbridge *et al*, 2006; van Schijndel *et al*, 2011). The remaining candidate SNPs do not cause missense mutations, but have been discovered in genome-wide association and genetic risk studies of neuropsychiatric disease (Lambert *et al*, 2009; Silberberg *et al*, 2006; Konrad *et al*, 2009). To obtain genotype information, genomic DNA samples were analyzed on the Human610-Quad BeadChip (Illumina) according to the manufacturer's protocols (Infinium HD Assay; Super Protocol Guide; Rev. A, May 2008). Additionally, imputation was performed by mapping the genotyped

information to HapMap (Release 22 Build 36) using the Mach software (http://www.sph.umich.edu/csg/abecasis/MACH/index.html). All candidate SNPs passed a platform-specific quality control score (>0.7) and genotype call rate (>0.95).

D. Statistical Analysis

Linear mixed-effects models were used to study the joint and individual associations of genotypes with imaging measures, to take into account the relatedness between the subjects. For n subjects and p independent predictors (SNPs or other covariates), regression coefficients (β) were obtained, using the efficient mixed-model association (EMMA) software with restricted maximum likelihood estimation (Kang *et al*, 2008), according to the formula:

$$y = X\beta + Zb + \varepsilon$$

Here, y represents an n-component vector of voxelwise or mean FA measures, X is a matrix of SNP genotypes (coded additively as 0, 1 or 2 for the number of minor alleles) and/or covariates (e.g., sex and age), Z is the identity matrix, and b is a vector of random effects with a variance of $\sigma_g^2 K$, where K is the n by n kinship matrix for the twins and siblings (here, monozygotic twins are coded as 1, dizygotic twins and siblings as 0.5, and unrelated subjects as 0, corresponding to the expected proportion of their shared genetic polymorphisms, respectively). ε is a matrix of residual effects with a variance of $\sigma_e^2 I$, and I is an identity matrix. P-values for the significance of individual and joint SNP associations with FA were assessed using an F-test, according to the formula,

$$F = \frac{(RSS_{covariates} - RSS_{full})/(p_{full} - p_{covariates})}{RSS_{full}/(n - p_{full})}$$

where *RSS* represents the residual sum-of-squares, a reduced model includes only covariates, and a full model contains both SNPs and covariates. For all statistical analyses, the LONI pipeline

(http://pipeline.loni.ucla.edu/) was used for parallelization on a multi-CPU grid computer. The standard false discovery rate (FDR) method (Benjamini and Hochberg, 1995) was used for multiple comparison correction across voxels in the corpus callosum.

Results

We assessed six candidate SNPs that have recently been implicated, to varying degrees, in affecting the brain, at the gross anatomical or microstructural level. We first used linear mixed-effects models to regress each subject's genotype at each candidate SNP against average FA measures across the corpus callosum (all callosal voxels with an FA above 0.25), to study their individual effects on white matter integrity. The regression β- and p-values for each SNP on its own, treated as an independent predictor, are shown in **Table 7.1**.

Table 7.1. Associations of single SNPs with mean callosal FA. Candidate single nucleotide polymorphisms (SNPs) are listed with corresponding gene names and chromosome numbers. Minor allele frequencies (MAFs) were estimated from unrelated subjects in our dataset, for each genetic variant; the resulting MAF estimates were, in all cases, comparable to corresponding MAF estimates reported in the literature. *P*-values arising from tests of deviation from Hardy-Weinberg Equilibrium (HWE) are also reported for each SNP, none of which were significant. Linear regression coefficients, β, are also shown (with directions of effect corresponding to the number of *minor* alleles). *P*-values report the significance of association of each SNP with the average fractional anisotropy (FA) across the corpus callosum (CC). The top three genes have SNPs that are each predictive of CC integrity on their own, while the others individually do not, at least in this sample.

SNP	Gene	Chromosome	MAF	HWE *p*-value	*B* (*p*-value)
rs6336	*NTRK1*	1	0.04	0.062	-0.015 (**0.0066**)
rs11136000	*CLU*	8	0.45	0.072	0.0051 (**0.037**)
rs839523	*ErbB4*	2	0.30	0.42	0.0053 (**0.036**)
rs4680	*COMT*	22	0.48	0.50	0.0039 (0.11)
rs1799945	*HFE*	6	0.16	0.10	0.0032 (0.29)
rs6265	*BDNF*	11	0.21	0.085	0.00031 (0.91)

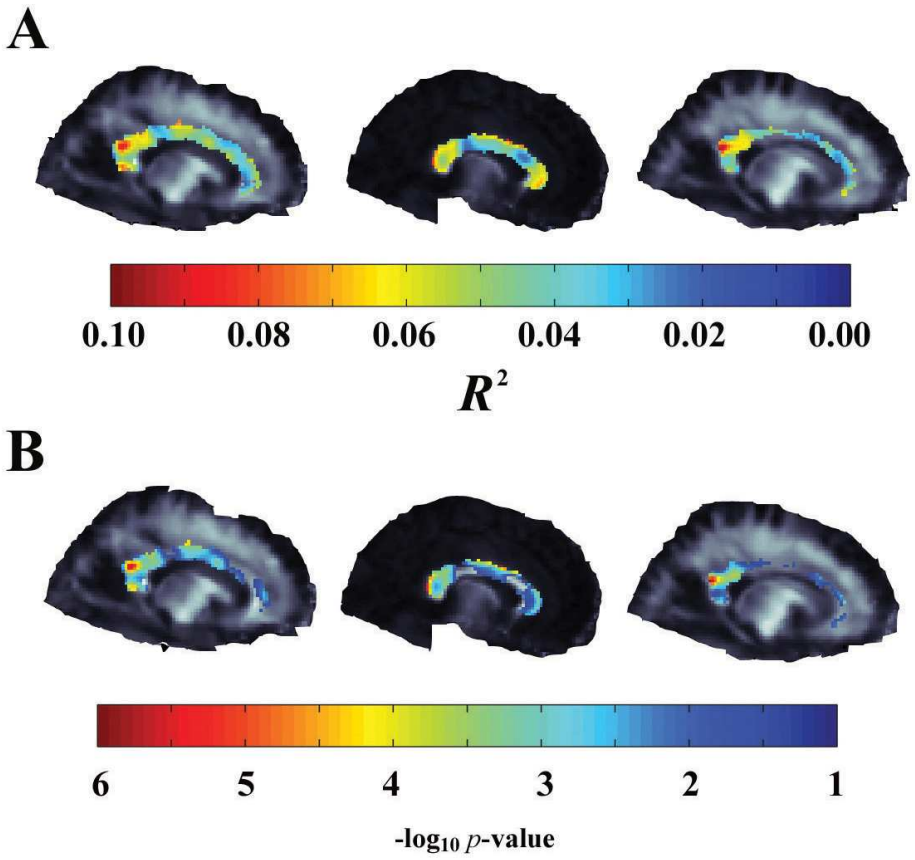
We then assessed the joint effect of our set of candidate SNPs on the corpus callosum, using a partial *F*-test and linear mixed-effects model to compute *p*-values. When regressed against average FA across the corpus callosum, a 5-SNP model with *NTRK1, CLU, COMT, ErbB4* and *HFE* containing top SNPs explained 5.6% of the variance in FA (p=0.0001; model included sex and age; a model including only sex and age explained 0.42% of the variance). Prediction of mean callosal FA was improved by adding candidate SNPs in a stepwise fashion (**Table 7.2**). Addition of the *BDNF* SNP, however, did not improve the model. To ensure multicollinearity was not present among the genotypes, we looked the correlation structure between the candidate SNPs, and none was correlated with any of the others (**Table 7.3**).

Table 7.2. Multilocus effects on average callosal FA and 3-dimensional maps of the corpus callosum. Groups of candidate single nucleotide polymorphisms (SNPs) are considered in joint association with mean callosal FA and voxelwise FA measures in the corpus callosum, in a stepwise fashion, adding SNPs in order of their individual effects (strongest to weakest). For mean FA, the fractions of variability explained by the top SNPs are shown, using the multiple regression R^2. For voxelwise results**,** we show the critical *p*-value thresholds after correcting for multiple comparisons with the false discovery rate (FDR) method, the number of voxels passing FDR at the critical thresholds (also as a percentage of all callosal voxels), the number of voxels with joint effect *p*-values less than p=0.05 (also as a percentage of all callosal voxels) and the minimum joint effect *p*-values across the corpus callosum. Values are displayed for each progressively expanding set of SNP predictors. For all associations, we adjusted for sex and age at each voxel; we accounted for kinship structure via mixed-effects models.

	Mean callosal FA	**3D maps of the corpus callosum**			
No. of top SNPs	R^2	**critical FDR *p***	**voxels below critical *p***	**voxels with $p<0.05$**	**minimum *p***
1	0.031	3.49×10^{-2}	4303 (70%)	4709 (77%)	1.49×10^{-5}
2	0.039	4.02×10^{-2}	4946 (80%)	5147 (84%)	8.54×10^{-7}
3	0.046	3.96×10^{-2}	4874 (79%)	5080 (83%)	1.23×10^{-6}
4	0.052	4.06×10^{-2}	5003 (81%)	5164 (84%)	1.54×10^{-6}
5	0.056	4.08×10^{-2}	5024 (82%)	5184 (84%)	1.25×10^{-6}
6	0.056	3.66×10^{-2}	4501 (73%)	4830 (79%)	4.25×10^{-6}

We also investigated the combined influence of the candidate SNPs on more detailed, spatial maps of the corpus callosum. In a stepwise fashion, in order of the SNPs' individual effects (strongest to weakest, as shown in **Table 7.1**), we studied multilocus effects on voxel-by-voxel maps of callosal white matter structure are shown in **Table 7.2**. The 5-SNP model showed the most widespread, statistically significant influence on the corpus callosum, where 82% of voxels (encompassing the callosal body, *genu* and *splenium*) survived the false discovery rate (FDR) correction for multiple comparisons across all callosal voxels, at a critical *p*-value threshold of 0.041 (**Figure 7.2**); we note that in FDR, a *higher* critical *p*-value denotes a stronger effect, as it is the highest threshold that controls the FDR; as such this effect is widespread and strong. Both the number of statistically significant voxels and the critical *p*-value threshold were strongest for the 5-SNP model. Figure 2 also shows the voxelwise distribution of the fraction of variance explained by the 5-SNP model in the corpus callosum.

Figure 7.2. Voxelwise R^2 and *p*-values are shown in three representative sagittal slices for the joint effect of 5 SNPs in the *NTRK1, CLU, COMT, ErbB4* and *HFE* genes on the corpus callosum microstructure, measured by DTI fractional anisotropy (FA). (A) The coefficient of determination (R^2) or predictability of the 5-SNP model at each voxel is shown in the selected slices. Warmer colors represent higher fractions of variance in FA explained by the multi-SNP model. (B) *P*-values are shown for the 5-SNP model at each voxel; maps are corrected for multiple comparisons across voxels by applying a critical *p*-value threshold to control the FDR. Warmer colors represent more significant associations (greater effect sizes). For associations at each voxel, we adjusted for any effects of sex and age, and accounted for kinship structure via mixed-effects models.



In addition to the additive, linear model, we included two-way SNP-SNP interactions in the mean callosal FA mixed-effect model. No significant interactions were found (**Table 7.4**). We also explored prediction of voxel-by-voxel FA from the 5 SNPs using two popular machine learning models (support vector regression and artificial neural networks; see Supplementary Methods), within a cross-validation framework, which similarly to the mixed-effect model, led to

statistically significant predictions across the corpus callosum (**Figure 7.3**). At each voxel, mean-squared errors of predictions of FA were obtained from the candidate genotypes. The artificial neural network and support vector regression models' predictive errors were then compared to those of null predictors (i.e., where FA is randomly assigned to subjects) through permutations. The learning models were found to be statistically significant across 75% and 40% of the corpus callosum voxels, after correcting for multiple comparisons, with critical *p*-value thresholds of 0.037 and 0.019, respectively.

Discussion

In this work, we aimed to predict neuroanatomical white matter structure based on multiple genetic risk factors, while covarying for sex and age. Five of the six candidate polymorphisms that we considered in the study - *CLU* (Braskie *et al*, 2011b), *ErbB4* (Konrad *et al*, 2009), *NTRK1* (Braskie *et al*, 2011a), *COMT* (Thomason *et al*, 2010) and *HFE* (Jahanshad *et al*, 2012a) – explained close to 6% of the variability in mean callosal FA, using a linear mixed-effect model. This is a considerable fraction of the variance explained by only a few SNPs, taking into account the complexity of the structure and the non-genetic factors that influence it. It is also comparable to previous findings in the literature for multilocus models of a brain imaging phenotype. Biffi *et al* (2010) found that 3% of the variance in MRI-derived volumes of several brain regions could be explained from a number of candidate genes for Alzheimer's disease. Nikolova *et al* (2011) showed 11% of the variance in ventral striatal reactivity could be explained from their panel of five polymorphisms. We also found that our candidate polymorphisms displayed extensive, significant effects on 82% of the volume of the corpus callosum, when cumulatively modeled at a voxelwise FA basis, which captures more spatial detail than an average measure of FA across

the corpus callosum. We also confirmed significant predictions across the corpus callosum from the 5 SNPs using multilocus machine learning models. These yielded similar predictions, but were less spatially widespread, as only a subset of subjects were considered who were unrelated to each other.

We focused on DTI-derived fractional anisotropy (FA) of the corpus callosum as our imaging measure in this study. The corpus callosum is the largest white matter structure in the brain, containing over 300 million axons (Hofer and Frahm, 2006). This fiber bundle transfers motor, sensory and cognitive information between the two cerebral hemispheres (Huang *et al*, 2005). With the advent of diffusion tensor imaging, it has become increasingly clear that the structure of the corpus callosum is impaired in several brain disorders. In a recent meta-analysis, for instance, Patel *et al* (2011) concluded that the *splenium* of the corpus callosum has significantly lower FA in patients with schizophrenia versus controls. Recent DTI studies have also identified callosal abnormalities in patients with other brain disorders such as bipolar disorder (Benedetti *et al*, 2011), post-traumatic stress disorder (Jackowski *et al*, 2008) and autism (Alexander *et al*, 2007). It would therefore be beneficial, clinically, to know an individual's personalized genetic risk for a corpus callosum structural abnormality. In addition, the microstructure of the corpus callosum has been shown to be highly heritable in studies including those with the same Australian twins as in this paper. Chiang *et al* (2009) mapped out genetic contributions to white matter structure in the Australian twins and discovered significant voxelwise effects in the callosal *genu* and *splenium*. In that paper, a classical twin design was used to estimate the overall genetic contribution to the observed variance, but effects of specific SNPs were not assessed or modeled. Similarly, Kochunov *et al* (2010) found mean FA values from the *corpus*, body and *genu* of the

corpus callosum were highly heritable (all with $h^2 > 0.5$) in members of the San Antonio Family Study. Similar results have also been reported in studies of young children (Brouwer *et al*, 2010) and in older individuals (e.g., Pfefferbaum *et al*, 2001). Recently, it was also shown in the same Australian population as ours that the heritability of callosal FA, particularly in the *genu*, is high regardless of imaging protocol differences (Jahanshad *et al*, 2012b). Here, we show that predictions of microstructural measures may be made based on a few common polymorphisms. We focused on the corpus callosum here, but our results may also have implications for other white matter tracts in the brain. The millions of axons in the corpus callosum connect numerous regions of the brain with each other. Genetic variants that affect this brain structure may also have roles in other white matter regions.

We selected six candidate SNPs for our study based on their reported individual effects on white matter structure on DTI and their importance in neuropsychiatric disease. The val158met missense mutation resulting from the candidate SNP in *COMT* causes reduced degradation and thus increased availability of dopamine, thereby leading to alterations in reward experience, executive function and working memory, with implications on risk for neuropsychiatric disease and differential response to therapy (Tunbridge *et al*, 2006; Wichers *et al*, 2008). *BDNF*'s val66met polymorphism, which affects the neurotrophin's secretion and its function in long-term potentiation, has been investigated in many studies and shown to alter memory performance at a young age, among other associations with neuropsychiatric disease (Egan *et al*, 2003; Hariri *et al*, 2003). Similarly, although not as fully characterized, the candidate SNP in the neurotrophin receptor gene, *NTRK1*, leads to a his598tyr amino acid change in the kinase domain of TrkA, and has been significantly associated with risk for schizophrenia (van Schijndel *et al*, 2009; van

Schijndel *et al*, 2011). In our study, this SNP had the strongest effect of all candidates on white matter structure. We also found subjects with greater numbers of minor alleles of the *NTRK1* polymorphism had lower FA, which is consistent with data suggesting that the minor allele is over-represented in schizophrenia patients (van Schijndel *et al*, 2011). Another receptor gene we considered was the neuregulin receptor, *ErbB4*, with an intronic variant associated with schizophrenia risk in several studies (Konrad *et al*, 2009; Nicodemus *et al*, 2006, Silderberg *et al*, 2006). Another intronic variant, rs11136000, in *CLU* has been discovered and replicated in genome-wide association studies of Alzheimer's disease (Lambert *et al*, 2009). Similarly, the his63asp mutation in iron-related *HFE* gene has been linked to Alzheimer's disease, along with other neurodegenerative disorders (Connor and Lee, 2006). We found all variants except for the one in *BDNF* contributed additively to prediction of average callosal FA as well as three-dimensional maps of voxelwise FA across the corpus callosum.

Personalized prediction of individuals' disease-related measures is being advocated by some as a vital component of future diagnosis and treatment of brain disorders (Koslow *et al*, 2010). Some of the genetic variation may account for some of the broad heterogeneity in patients' disease status (Cummings, 2000; Folstein and Rosen-Sheidley, 2001) and the extent to which they respond to therapy (Gordon, 2007). Multilocus models are particularly appealing for personalized prediction of disease. Several groups have explored multilocus models of candidate risk variants in the context of brain disorders. Carayol *et al* (2010), for instance, reported on the cumulative effect of four candidate SNPs on the risk for autism, using a case-control approach. These models are beginning to be applied to brain imaging in the context of neuropsychiatric

disorders (Biffi *et al*, 2010; Hibar *et al*, 2011a; Nikolova *et al*, 2011), and may provide more biologically meaningful predictions with implications for personalized diagnosis and therapy.

Future studies are needed to replicate our findings in independent cohorts of subjects, even though we found significant predictions using cross-validation in the SVM and ANN analyses. In addition, as new candidate gene studies and genome-wide searches using DTI measures (e.g., Kochunov *et al*, 2011) identify effects of new variants, candidate genes may be added or removed from this panel, to better predict white matter integrity. We did not find evidence for two-way interactions between the SNPs in our study, which is probably reasonable, as the SNPs are likely contributing independently and additively to white matter integrity, and interactions are second-order effects (modulations of the main effect of a gene) that may require large samples to identify, if present at all. Such interactions, however, may be identified in follow-up studies particularly with SNPs that directly share the same pathway, like *NTRK2* and *BDNF* (Perroud *et al*, 2009), *NRG1* and *ErbB4* (Nicodemus *et al*, 2010) or *COMT* and *5-HTTLPR* (Borroni *et al*, 2006). In this paper, we took a voxelwise approach to study genetic associations with FA. In addition to voxelwise maps of FA, tract- and fiber-based measures from diffusion imaging may also be considered as predictive outputs. Such measures, along with multivariate methods that simultaneously consider not only multiple genes, but also multiple voxels (Vounou *et al*, 2010; Hibar *et al*, 2011b; Wan *et al*, 2011) may help provide more statistical power. For instance, our voxelwise, multilocus model improved only slightly beyond the 2-SNP model with polymorphisms in *NTRK1* and *CLU*. This may be due to the strong effects of *NTRK1* and *CLU* SNPs on their own, but it may also be because multiple variants do not necessarily affect the same exact voxels. This may make it difficult to obtain substantially more expansive voxelwise

effects by adding more variants to the model. Here, we considered genetic polymorphisms as predictors, and these explained a small but significant proportion of the heritable variation in white matter structure across young, healthy individuals. Although it remains to be determined, it is plausible that a measure of white matter integrity, such as DTI-derived FA, relates to a person's lifetime risk for developing mental and neurodegenerative disorders, especially for disorders in which FA is abnormally low.

Supplementary Information

Machine Learning

Artificial neural networks (ANNs) comprise a group of machine learning algorithms, in which input patterns, or nodes, are connected, weighted, and translated into output nodes via an activation function. ANNs are generally formulated as $y = f(\varphi(x, w))$, where y, x, and w represent the input, output and weights respectively, and φ typically represents a linear combination of the inputs. f is the activation function, typically a logistic function. We implemented a three-layer ANN, consisting of an input layer with SNPs and covariates, an output layer with DTI-derived FA measures, and a hidden layer, using the 'nnet' package (Venables and Ripley, 2002) in R (http://cran.r-project.org). We also included weight decay regularization in the ANN, which introduces an upper bound on the sum of the squares of the weights (Hinton 1986), and is related to the regularization used in penalized regression or ridge regression (Hoerl, 1962), and helps with the machine's generalizability (Jain *et al*, 2000).

Support vector regression (SVR) is another machine learning algorithm related to support vector machines (Vapnik, 1995). Training with patterns (SNP genotype profiles, along with sex and

age, in our case) and known outcomes (DTI-derived FA, in this study) leads to the construction of a hyperplane, which is then applied to predict outcomes in testing patterns. The traditional ε-SVR is formulated so that the differences between observed and predicted measures are to be no greater than the parameter ε. We implemented the newer ν-SVR algorithm within the 'e1071' package (Dimitriadou *et al*, 2005) in R. ν-SVR modifies the optimization formulation of ε-SVR through the ν parameter, which introduces a lower bound on the fraction of predictions allowed to deviate by more than ε from observed measures (Basak *et al*, 2007). The modified optimization problem is:

$$\min \frac{1}{2}\|w\|^2 + C\left(\nu\xi + \frac{1}{l}\sum(\xi_i + \xi_i^*)\right)$$

$$\text{subject to } ((w \cdot x_i) + b) - y_i \le \varepsilon + \xi_i$$
$$y_i - ((w \cdot x_i) + b) \le \varepsilon + \xi_i^*$$
$$\xi_i^* \ge 0$$
$$\varepsilon \ge 0$$

Statistical significance of predictions was obtained by running ANN and SVR processes 1,000 times, on permuted measures (the voxelwise FA measures or outputs were scrambled across the subjects), and leave-one-out predictive errors were obtained at every voxel for each permutation.

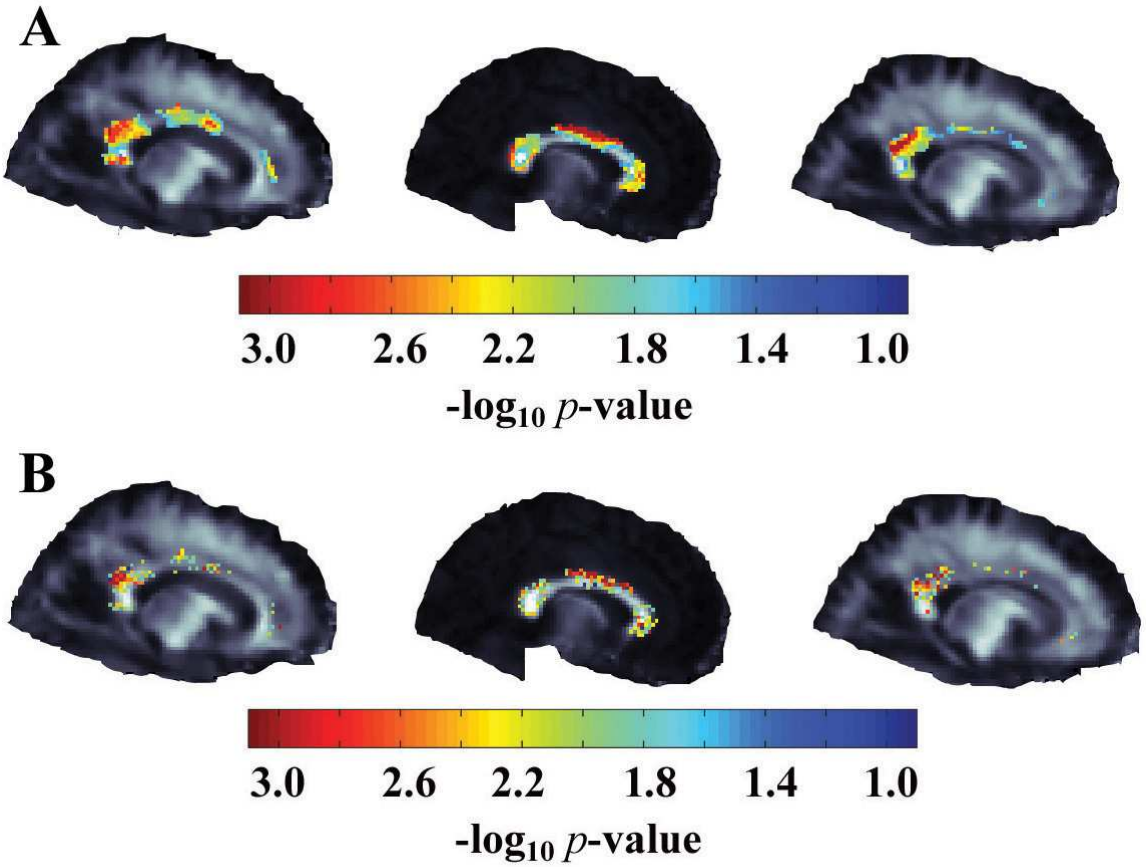
Table 7.3. Correlation (measured by r^2) between the six candidate single nucleotide polymorphisms is shown to ensure multicollinearity does not affect regression results. This analysis was performed on a subset of 246 subjects, not related to each other, to avoid bias due to subject kinship.

SNPs	rs11136000	rs6336	rs4680	rs839523	rs6265	rs1799945
rs11136000	1					
rs6336	5.74×10^{-3}	1				
rs4680	5.92×10^{-5}	4.39×10^{-4}	1			
rs839523	1.16×10^{-2}	2.42×10^{-3}	2.62×10^{-3}	1		
rs6265	3.76×10^{-3}	1.94×10^{-3}	1.58×10^{-2}	6.64×10^{-6}	1	
rs1799945	2.60×10^{-3}	1.95×10^{-3}	2.39×10^{-3}	1.83×10^{-3}	3.45×10^{-3}	1

Table 7.4. All possible two-way interactions between the candidate SNPs are included in a multiple linear mixed-effects model. Candidate SNPs are denoted by their corresponding gene. Individual SNP genotypes, sex and age are also included along with all interaction terms in the mixed-effects regression model. No interaction terms are significant after correction for multiple comparisons.

SNP x SNP	***β***	***P*-value**
CLU x *HFE*	4.28×10^{-3}	3.39×10^{-1}
CLU x *NTRK1*	6.90×10^{-4}	9.40×10^{-1}
CLU x *COMT*	-1.50×10^{-3}	6.79×10^{-1}
CLU x *ErbB4*	-7.33×10^{-3}	5.59×10^{-2}
CLU x *BDNF*	-1.37×10^{-3}	7.50×10^{-1}
HFE x *NTRK1*	-3.89×10^{-3}	7.94×10^{-1}
HFE x *COMT*	8.67×10^{-3}	4.94×10^{-2}
HFE x *ErbB4*	-6.43×10^{-3}	1.66×10^{-1}
HFE x *BDNF*	9.31×10^{-3}	3.70×10^{-2}
NTRK1 x *COMT*	9.52×10^{-3}	4.17×10^{-1}
NTRK1 x *ErbB4*	-4.21×10^{-3}	6.20×10^{-1}
NTRK1 x *BDNF*	2.29×10^{-4}	9.81×10^{-1}
COMT x *ErbB4*	4.40×10^{-3}	2.37×10^{-1}
COMT x *BDNF*	-3.66×10^{-3}	3.48×10^{-1}
ErbB4 x *BDNF*	-4.89×10^{-4}	9.08×10^{-1}

Figure 7.3. (A) Genotypes for 5 candidate SNPs in the *NTRK1, CLU, COMT, ErbB4* and *HFE* genes are incorporated into an artificial neural network (ANN) model to predict voxelwise FA measures. The ANN's mean squared error (MSE) was divided by the MSE for a null predictor that outputs an average FA for the voxel. Permutation tests were conducted at each voxel to assess the significance of the predictive model. Permutation-based *p*-values are shown for three sagittal FA slices. Warmer colors represent more statistically significant predictions. (B) Genotypes for the 5 candidate SNPs are incorporated into a support vector regression (SVR) model to predict voxelwise FA measures. The model is evaluated as in (A). For both (A) and (B), analyses are performed only in a subset of 246 subjects, who are not related to each other, in order to avoid their kinship as a confounding factor.



References

Alexander AL, Lee JE, Lazar M, Boudos R, DuBray MB, Oakes TR, Miller JN, Lu J, Jeong EK, McMahon WM, Bigler ED, Lainhart JE. Diffusion tensor imaging of the corpus callosum in Autism. NeuroImage. 2007; 34(1):61-73.

Allen SJ, Dawbarn D. Clinical relevance of the neurotrophins and their receptors. Clin Sci (Lond). 2006; 110(2):175-191.

Bartzokis G, Lu PH, Tingus K, Peters DG, Amar CP, Tishler TA, Finn JP, Villablanca P, Altshuler LL, Mintz J, Neely E, Connor JR. Gender and iron genes may modify associations between brain iron and memory in healthy aging. Neuropsychopharmacology. 2011; 36(7):1375-1384.

Basak D, Pal S, Patranabis DC. Support vector regression. Neural Inform Process Lett Rev. 2007; 11(10):203-224.

Benedetti F, Yeh PH, Bellani M, Radaelli D, Nicoletti MA, Poletti S, Falini A, Dallaspezia S, Colombo C, Scotti G, Smeraldi E, Soares JC, Brambilla P. Disruption of white matter integrity in bipolar depression as a possible structural marker of illness. Biol Psychiatry. 2011; 69():309-317.

Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc B. 1995; 57:289-300.

Biffi A, Anderson CD, Desikan RS, Sabuncu M, Cortellini L, Schmansky N, Salat D, Rosand J; Alzheimer's Disease Neuroimaging Initiative (ADNI). Genetic variation and neuroimaging measures in Alzheimer disease. Arch Neurol. 2010; 67(6):677-685.

Binder DK, Scharfman HE. Brain-derived neutrotrophic factor. Growth Factors. 2004; 22(3):123-131.

Borroni B, Grassi M, Agosti C, Archetti S, Costanzi C, Cornali C, Caltagirone C, Caimi L, Di Luca M, Padovani A. Cumulative effect of COMT and 5-HTTLPR polymorphisms and their interaction with disease severity and comorbidities on the risk of psychosis in Alzheimer disease. Am J Geriatr Psychiatry. 2006; 14(4):343-351.

Braskie MN, Jahanshad N, Stein JL, Barysheva M, Johnson K, McMahon KL, de Zubicaray GI, Martin NG, Wright MJ, Ringman JM, Toga AW, Thompson PM. Schizophrenia risk variant in the NTRK1 gene relates to lower white matter integrity in young healthy adults. J Neurosci. submitted. 2011a.

Braskie MN, Jahanshad N, Stein JL, Barysheva M, McMahon KL, de Zubicaray GI, Martin NG, Wright MJ, Ringman JM, Toga AW, Thompson PM. (2011b). Common Alzheimer's disease risk variant within the CLU gene affects white matter microstructure in young adults. J Neurosci. 31(18):6764-6770.

Brouwer RM, Mandl RC, Peper JS, van Baal GC, Kahn RS, Boomsma DI, Hulshoff Pol HE. Heritability of DTI and MTR in nine-year-old children. NeuroImage. 2010; 53():1085-1092.

Carayol J, Schellenberg GD, Tores F, Hager J, Ziegler A, Dawson G. Assessing the impact of a combined analysis of four common low-risk genetic variants on autism risk. Mol Autism. 2010; 1(1):4.

Chiang MC, Barysheva M, Shattuck DW, Lee AD, Madsen SK, Avedissian C, Klunder AD, Toga AW, McMahon KL, de Zubicaray GI, Wright MJ, Srivastava A, Balov N, Thompson PM. Genetics of brain fiber architecture and intellectual performance. J Neurosci. 29(7):2212-2224.

Chiang MC, Barysheva M, Toga AW, Medland SE, Hansell NK, James MR, McMahon KL, de Zubicaray GI, Martin NG, Wright MJ, Thompson PM. BDNF gene effects on brain circuitry replicated in 455 twins. NeuroImage. 2011a; 55(2):448-454.

Chiang MC, McMahon KL, de Zubicaray GI, Martin NG, Hickie I, Toga AW, Wright MJ, Thompson PM. Genetics of white matter development: a DTI study of 705 twins and their siblings aged 12 to 29. NeuroImage. 2011b; 54(3):2308-2317.

Connor JR, Lee SY. HFE mutations and Alzheimer's disease. J Alzheimer Dis. 2006; 10(2-3):267-276.

Cummings JL. Cognitive and behavioral heterogeneity in Alzheimer's disease: seeking the neurobiological basis. Neurobiol Aging. 2000; 21(6):845-861.
de Geus E, Goldberg T, Boomsma DI, Posthuma D. Imaging the genetics of brain structure and function. Biol Psychol. 2008; 79(1):1-8.
de Zubicaray GI, Chiang MC, McMahon KL, Shattuck DW, Toga AW, Martin NG, Wright MJ, Thompson PM. Meeting the Challenges of Neuroimaging Genetics. Brain Imaging Behav. 2008; 2(4):258-263.
DeMattos RB, O'dell MA, Parsadanian M, Taylor JW, Harmony JA, Bales KR, Paul SM, Aronow BJ, Holtzman DM. Clusterin promotes amyloid plaque formation and is critical for neuritic toxicity in a mouse model of Alzheimer's disease. Proc Natl Acad Sci USA. 2002; 99(16):10843-10848.
Dimitriadou E, Hornik K, Leisch F, Meyer D, Weingessel A. Misc Functions of the Department of Statistics (e1071), *TU Wien*. 2005.
Egan MF, Weinberger DR, Lu B. Schizophrenia, III: brain-derived neurotropic factor and genetic risk. Am J Psychiatry. 2003; 160(7):1242.
Ellison-Wright I, Bullmore E. Meta-analysis of diffusion tensor imaging studies in schizophrenia. Schizophr Res. 2009; 108(1-3):3-10.
Fellgiebel A, Wille P, Müller MJ, Winterer G, Scheurich A, Vucurevic G, Schmidt LG, Stoeter P. Ultrastructural hippocampal and white matter alterations in mild cognitive impairment: a diffusion tensor imaging study. Dement Geriatr Cogn Disord. 2004; 18(1):101-108.
Folstein SE, Rosen-Sheidley B. Genetics of autism: complex aetiology for a heterogeneous disorder. Nat Rev Genet. 2001; 2(12):943-955.
Foong J, Maier M, Clark CA, Barker GJ, Miller DH, Ron MA. Neuropathological abnormalities of the corpus callosum in schizophrenia: a diffusion tensor imaging study. J Neurol Neurosurg Psychiatry. 2000; 68(2):242-244.
Gordon E. Integrating genomics and neuromarkers for the era of brain-related personalized medicine. Personalized Medicine. 2007; 4:201-215.
Hahn CG, Wang HY, Cho DS, Talbot K, Gur RE, Berrettini WH, Bakshi K, Kamins J, Borgmann-Winter KE, Siegel SJ, Gallop RJ, Arnold SE. Altered neuregulin 1-erbB4 signaling contributes to NMDA receptor hypofunction in schizophrenia. Nat Med. 2006; 12(7):824-828.
Hall MH, Smoller JW. A new role for endophenotypes in the GWAS era: functional characterization of risk variants. Harv Rev Psychiatry. 2010; 18(1):67-74.
Hariri AR, Goldberg TE, Mattay VS, Kolachana BS, Callicott JH, Egan MF, Weinberger DR. Brain-derived neurotrophic factor val66met polymorphism affects human memory-related hippocampal activity and predicts memory performance. J Neurosci. 2003; 23(17):6690-6694.
Heng S, Song AW, Sim K. White matter abnormalities in bipolar disorder: insights from diffusion tensor imaging studies. J Neural Transm. 2010; 117(5):639-654.
Hibar DP, Stein JL, Kohannim O, Jahanshad N, Saykin AJ, Shen L, Kim S, Pankratz N, Foroud T, Huentelman MJ, Potkin SG, Jack CR Jr, Weiner MW, Toga AW, Thompson PM; Alzheimer's Disease Neuroimaging Initiative. Voxelwise gene-wide association study (vGeneWAS): multivariate gene-based association testing in 731 elderly subjects. NeuroImage. 2011a; 56(4):1875-1891.
Hibar DP, Kohannim O, Stein JL, Chiang MC, Thompson PM. Multilocus genetic analysis of brain images. Front Genet. 2011b; 2:73. doi: 10.3389/fgene.2011.00073.

Hinton GE. Learning distributed representations of concepts. In: Proceedings of the Eighth Annual Meeting of the Cognitive Science Society. Hillsdale: New Jersey. pp 1-12. 1986.
Hoerl AE. Application of ridge analysis to regression problems. Chem Eng Prog. 1962; 58: 54-59.
Hofer S, Frahm J. Topography of the human corpus callosum revisited--comprehensive fiber tractography using diffusion tensor magnetic resonance imaging. NeuroImage. 2006; 32(3):989-994.
Huang H, Zhang J, Jiang H, Wakana S, Poetscher L, Miller MI, van Zijl PC, Hillis AE, Wytik R, Mori S. DTI tractography based parcellation of white matter: application to the mid-sagittal morphology of corpus callosum. NeuroImage. 2005; 26(1):195-205.
Jackowski AP, Douglas-Palumberi H, Jackowski M, Win L, Schultz RT, Staib LW, Krystal JH, Kaufman J. Corpus callosum in maltreated children with posttraumatic stress disorder: a diffusion tensor imaging study. Psychiatry Res. 2008; 162(3):256-261.
Jahanshad N, Kohannim O, Hibar DP, Stein JL, McMahon KL, de Zubicaray GI, Medland SE, Montgomery GW, Whitfield JB, Martin NG, Wright MJ, Toga AW, Thompson PM. Brain structure in healthy adults is related to serum transferrin and the H63D polymorphism in the HFE gene. Proc Natl Acad Sci USA. 2012a; doi: 10.1073/pnas.1105543109.
Jahanshad N, Kohannim O, Toga AW, McMahon KL, de Zubicaray GI, Martin NG, Wright MJ, Thompson PM. Diffusion imaging protocol effects on genetic associations. International Symposium on Biomedical Imaging (ISBI). Barcelona, Spain. 2012b.
Jahanshad N, Lee AD, Barysheva M, McMahon KL, de Zubicaray GI, Martin NG, Wright MJ, Toga AW, Thompson PM. Genetic influences on brain asymmetry: a DTI study of 374 twins and siblings. NeuroImage. 2010; 52(2):455-469.
Jain AK, Duin RPW, Mao J. Statistical pattern recognition: a review. IEEE Trans. Pattern Anal. Machine Intell. 2000; 22: 4–37.
Jones DK, Horsfield MA, Simmons A. Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. Magn Reson Med. 1999; 42(3):515-525.
Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Eskin E. Efficient control of population structure in model organism association mapping. Genetics 2008; 178(3):1709-1723.
Ke X, Tang T, Hong S, Hang Y, Zou B, Li H, Zhou Z, Ruan Z, Lu Z, Tao G, Liu Y. White matter impairments in autism, evidence from voxel-based morphometry and diffusion tensor imaging. Brain Res. 2009; 1265:171-177.
Kochunov P, Glahn DC, Lancaster JL, Winkler AM, Smith S, Thompson PM, Almasy L, Duggirala R, Fox PT, Blangero J. Genetics of microstructure of cerebral white matter using diffusion tensor imaging. NeuroImage 2010; 53(3):1109-1116.
Kochunov P, Glahn DC, Nichols TE, Winkler AM, Hong EL, Holcomb HH, Stein JL, Thompson PM, Curran JE, Carless MA, Olvera RL, Johnson MP, Cole SA, Kochunov V, Kent J, Blangero J. Genetic analysis of cortical thickness and fractional anisotropy of water diffusion in the brain. Front Neurosci. 2011; 5:120.
Konrad A, Vucurevic G, Musso F, Stoeter P, Dahmen N, Winterer G. ErbB4 genotype predicts left frontotemporal structural connectivity in human brain. Neuropsychopharmacology. 2009; 34(3):641-650.
Konrad K, Eickhoff SB. Is the ADHD brain wired differently? A review on structural and functional connectivity in attention deficit hyperactivity disorder. Hum Brain Mapp. 2010; 31(6):904-916.

Koslow SH, Williams LM, Gordon E. Personalized medicine for the brain: a call for action. Mol Psychiatry. 2010; 15(3):229-230.

Lambert JC, Heath S, Even G, Campion D, Sleegers K, Hiltunen M, Combarros O, Zelenika D, Bullido MJ, Tavernier B, Letenneur L, Bettens K, Berr C, Pasquier F, Fiévet N, Barberger-Gateau P, Engelborghs S, De Deyn P, Mateo I, Franck A, Helisalmi S, Porcellini E, Hanon O; European Alzheimer's Disease Initiative Investigators, de Pancorbo MM, Lendon C, Dufouil C, Jaillard C, Leveillard T, Alvarez V, Bosco P, Mancuso M, Panza F, Nacmias B, Bossù P, Piccardi P, Annoni G, Seripa D, Galimberti D, Hannequin D, Licastro F, Soininen H, Ritchie K, Blanché H, Dartigues JF, Tzourio C, Gut I, Van Broeckhoven C, Alpérovitch A, Lathrop M, Amouyel P. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. Nat Genet. 2009; 41(10):1094-1099.

Lee AD, Lepore N, Barysheva M, Chou, Y-Y, Brun CC, Madsen SK, McMahon K, de Zubicaray GI, Wright MJ, Toga AW, Thompson PM. Gene Effects Mapped Using Fractional and Geodesic Anisotropy in Diffusion Tensor Images of 92 Monozygotic and Dizygotic Twins. In: Alexander D, Gee J, Whitaker R (eds.) MICCAI workshop on computational diffusion MRI. New York. pp 31-40. 2008.

Lee AD, Lepore N, de Leeuw J, Brun CC, Barysheva M, McMahon KL, de Zubicaray GI, Martin NG, Wright MJ, Thompson PM. Multivariate Variance-Components Analysis in DTI. In: ISBI '10. IEEE International Symposium on Biomedical Imaging. Rotterdam. pp 157-1160. 2010.

Leow A, Huang SC, Geng A, Becker J, Davis S, Toga A, Thompson P. Inverse consistent mapping in 3D deformable image registration: its construction and statistical properties. Inf Process Med Imaging. 2005; 19:493-503.

Marenco S, Radulescu E. Imaging genetics of structural brain connectivity and neural integrity markers. NeuroImage. 2010; 53(3):848-856.

Medland SE, Nyholt DR, Painter JN, McEvoy BP, McRae AF, Zhu G, Gordon SD, Ferreira MA, Wright MJ, Henders AK, Campbell MJ, Duffy DL, Hansell NK, Macgregor S, Slutske WS, Heath AC, Montgomery GW, Martin NG. Common Variants in the Trichohyalin Gene Are Associated with Straight Hair in Europeans. Am J Hum Genet. 2009; 85(5):750-755.

Meyer-Lindenberg A, Weinberger DR. Intermediate phenotypes and genetic mechanisms of psychiatric disorders. Nat Rev Neurosci. 2006; 7(10):818-827.

Mori S, Oishi K, Jiang H, Jiang L, Li X, Akhter K, Hua K, Faria AV, Mahmood A, Woods R, Toga AW, Pike GB, Neto PR, Evans A, Zhang J, Huang H, Miller MI, van Zijl P, Mazziotta J. Stereotaxic white matter atlas based on diffusion tensor imaging in an ICBM template. NeuroImage. 2008; 40(2):570-582.

Naggara O, Oppenheim C, Rieu D, Raoux N, Rodrigo S, Dalla Barba G, Meder JF. Diffusion tensor imaging in early Alzheimer's disease. Psychiatry Res. 2006; 146(3):243-249.

Nicodemus KK, Law AJ, Radulescu E, Luna A, Kolachana B, Vakkalanka R, Rujescu D, Giegling I, Straub RE, McGee K, Gold B, Dean M, Muglia P, Callicott JH, Tan HY, Weinberger DR. Biological validation of increased schizophrenia risk with NRG1, ERBB4, and AKT1 epistasis via functional neuroimaging in healthy controls. Arch Gen Psychiatry. 2010; 67(10):991-1001.

Nicodemus KK, Luna A, Vakkalanka R, Goldberg T, Egan M, Straub RE, Weinberger DR. Further evidence for association between ErbB4 and schizophrenia and influence on cognitive intermediate phenotypes in healthy controls. Mol Psychiatry. 2006; 11(12):1062-1065.

Nikolova YS, Ferrell RE, Manuck SB, Hariri AR (2011). Multilocus genetic profile for dopamine signaling predicts ventral striatum reactivity. Neuropsychopharmacology. 36(9):1940-1947.

Oishi K, Mielke MM, Albert M, Kyketsos CG. DTI Analyses and Clinical Applications in Alzheimer's Disease. J Alzheimers Dis. 2011; 26(Suppl):287-296.

Patel S, Mahon K, Wellington R, Zhang J, Chaplin W, Szeszko PR. A meta-analysis of diffusion tensor imaging studies of the corpus callosum in schizophrenia. Schizophr Res. 2011; 129(2-3):149-155.

Patel V, Chiang M-C, Thompson PM, McMahon KL, de Zubicaray GI, Martin NG, Wright MJ, Toga AW. Scalar Connectivity Measures from Fast Marching Tractography Reveal Heritability of White Matter Architecture. In: ISBI '10. IEEE International Symposium on Biomedical Imaging. Rotterdam. pp 1109-1112. 2010.

Perroud N, Aitchison KJ, Uher R, Smith R, Huezo-Diaz P, Marusic A, Maier W, Mors O, Placentino A, Henigsberg N, Rietschel M, Hauser J, Souery D, Kapelski P, Bonvicini C, Zobel A, Jorgensen L, Petrovic A, Kalember P, Schulze TG, Gupta B, Gray J, Lewis CM, Farmer AE, McGuffin P, Craig I. Genetic predictors of increase in suicidal ideation during antidepressant treatment in the GENDEP project. Neuropsychopharmacology. 2009; 34(12):2517-2528.

Pfefferbaum A, Sullivan EV, Carmelli D. Genetic regulation of regional microstructure of the corpus callosum in late life. Neuroreport. 2001; 12(8):1677-1681.

Silberberg G, Darvasi A, Pinkas-Kramarski R, Navon R. The involvement of ErbB4 with schizophrenia: association and expression studies. Am J Med Genet B Neuropsychiatr Genet. 2006; 141B(2):142-148.

Sussmann JE, Lymer GK, McKirdy J, Moorhead TW, Muñoz Maniega S, Job D, Hall J, Bastin ME, Johnstone EC, Lawrie SM, McIntosh AM. White matter abnormalities in bipolar disorder and schizophrenia detected using diffusion tensor magnetic resonance imaging. Bipolar Disord. 2009; 11(1):11-18.

Thomason ME, Dougherty RF, Colich NL, Perry LM, Rykhlevskaia EI, Louro HM, Hallmayer JF, Waugh CE, Bammer R, Glover GH, Gotlib IH. COMT genotype affects prefrontal white matter pathways in children and adolescents. NeuroImage 53(3):926-934.

Thomason ME, Thompson PM. Diffusion imaging, white matter, and psychopathology. Annu Rev Clin Psychol. 2011; 7:63-85.

Tunbridge EM, Harrison PJ, Weinberger DR. Catechol-o-methyltransferase, cognition, and psychosis: Val158Met and beyond. Biol Psychiatry. 2006; 60(2):141-151.

van Schijndel JE, van Loo KM, van Zweeden M, Djurovic S, Andreassen OA, Hansen T, Werge T, Kallunki P, Pedersen JT, Martens GJ. Three-cohort targeted gene screening reveals a non-synonymous TRKA polymorphism associated with schizophrenia. J Psychiatr Res. 2009; 43(15):1195-1199.

Van Schijndel JE, Van Zweeden M, Van Loo KM, Djurovic S, Andreassen OA, Hansen T, Werge T, Nyegaard M, Sørensen KM, Nordentoft M, Mortensen PB, Mors O, Børglum AD, Del-Favero J, Norrback KF, Adolfsson R, De Hert M, Claes S, Cichon S, Rietschel M, Nöthen MM, Kallunki P, Pedersen JT, Martens GJ. Dual association of a TRKA polymorphism with schizophrenia. Psychiatr Genet. 2011; 21(3):125-131.

Vapnik V. The Nature of Statistical Learning Theory. Springer, New York. 1995.

Venables WN, Ripley BD. Modern Applied Statistics with S. Springer, New York. 2002.

Versace A, Almeida JR, Hassel S, Walsh ND, Novelli M, Klein CR, Kupfer DJ, Phillips ML. Elevated left and reduced right orbitomedial prefrontal fractional anisotropy in adults with bipolar disorder revealed by tract-based spatial statistics. Arch Gen Psychiatry. 2008; 65(9):1041-1052.

Vounou M, Nichols TE, Montana G, for the Alzheimer's Disease Neuroimaging Initiative. Discovering genetic associations with high-dimensional neuroimaging phenotypes: A sparse reduced-rank regression approach. NeuroImage. 2010; 53(3):1147-1159.

Wan J, Kim S, Inlow M, Nho K, Swaminathan S, Risacheri SL, Fang S, Weiner MW, Beg MF, Wang L, Saykin AJ, Shen L; Alzheimer's Disease Neuroimaging Initiative. Hippocampal surface mapping of genetic risk factors in AD via sparse learning models. Med Image Comput Comput Assist Interv. 2011; 14(Pt 2):376-383.

White T, Nelson M, Lim KO. Diffusion tensor imaging in psychiatric disorders. Top Magn Reson Imaging. 2008; 19(2):97-109.

Wichers M, Aguilera M, Kenis G, Krabbendam L, Myin-Germeys I, Jacobs N, Peeters F, Derom C, Vlietinck R, Mengelers R, Delespaul P, van Os J. The catechol-O-methyl transferase Val158Met polymorphism and experience of reward in the flow of daily life. Neuropsychopharmacology. 2008; 33(13):3030-3036.

Yoo SY, Jang JH, Shin YW, Kim DJ, Park HJ, Moon WJ, Chung EC, Lee JM, Kim IY, Kim SI, Kwon JS. White matter abnormalities in drug-naive patients with obsessive-compulsive disorder: a diffusion tensor study before and after citalopram treatment. Acta Psychiatr Scand. 2007; 116(3):211-219.

Zecca L, Youdim MB, Riederer P, Connor JR, Crichton RR. Iron, brain ageing and neurodegenerative disorders. Nat Rev Neurosci. 2004; 5(11):863-873.

Chapter 8. Conclusions and future directions

For the first aim of my dissertation research, I pursued an automated classification algorithm, which would combine information from various biomarkers, including neuroimaging, biochemistry and genetics. As discussed in Chapter 3, we showed that such an approach based on machine learning can successfully integrate multiple biomarker modalities (e.g., magnetic resonance imaging, positron emission tomography, cerebrospinal fluid markers and genotyping) and provide reasonably high accuracies in diagnostic classification. In this particular case, we considered an Alzheimer's disease classifier, although this approach can be generalized to other disorders as well. In addition to accurate classification, we also showed that our classifier can help identify which biomarkers are more useful for which stages of disease. Lastly, we demonstrated that our multi-biomarker classification approach can be useful in selection of subjects for clinical trials. Fractions of subjects predicted by the automated classifier to be more advanced in the disease may be more likely to respond to treatment. This approach can therefore potentially reduce sample size requirements for trials. Future studies may find similar approaches useful in other disorders, where biomarkers specific to the disease and disease trajectory may be combined. More sophisticated machine learning algorithms may also be pursued, which can more powerfully combine diverse sources of biomarker data. Integration of information from novel biomarker such as new brain imaging modalities, high-throughput genotyping, gene expression and proteomics may also be of considerable value for Alzheimer's disease and other brain disorders. We and our collaborators have been pursuing several of these directions in ongoing, follow-up studies.

For the second aim of my graduate research, I explored new multilocus techniques to apply to the field of neuroimaging genetics. As reviewed in Chapter 4, these techniques consider multiple genetic variants jointly in association with images and imaging measures. Since they are often designed to better handle high-dimensional data and the interdependence between data elements (e.g., genetic variants), these techniques may provide more power than their alternative, univariate methods. As discussed in Chapter 5, we utilized a penalized regression method to explore the multilocus association of genomic regions with neuroimaging measures. We showed that several genetic variants were detected with higher statistical power, demonstrating the method's applicability in gene discovery. As shown in Chapter 6, we took another penalized regression approach to prediction of a neuroimaging measure from thousands of genotypes. We demonstrated that we can make signification predictions with this approach and also identify top genetic variants that contribute most to the predictions. Our multilocus neuroimaging genetic approaches and similar methods may be applied to other measures derived from neuroimaging in the future. We have been investigating other penalized regression methods in ongoing studies. Future research is needed to compare the statistical power of various penalized regression and related algorithms in the context of neuroimaging genetics. As reviewed in Chapter 4, there are already multivariate methods that simultaneously consider not only genotypes, but also whole images at a voxel-by-voxel basis. Application of these methods to new images and disorders may lead to exciting genetic discoveries. Meta-analysis neuroimaging genetics approaches such as ENIGMA (mentioned in Chapters 4-6) are also invaluable in future gene discovery and exploration, as they provide high statistical power due to the uniquely large numbers of subjects with imaging and genetic data.

For the third aim of my dissertation research, I pursued an automated predictor of brain integrity from genotype profiles. As discussed in Chapter 7, multiple common genetic variants were considered as neuropsychiatric risk candidates for the prediction of white matter integrity on DTI. We showed that several of our candidate single nucleotide polymorphisms jointly led to statistically significant predictions of fractional anisotropy on DTI. This was illustrated in maps of white matter microstructure as well as average fractional anisotropy across the corpus callosum. We explored several algorithms (mixed-effects multiple regression, support vector regression and artificial neural networks), all of which led to statistically significant predictions of brain integrity from the genotype signatures. Our approach has potential implications for early risk assessment of neuropsychiatric disorders, many of which cause white matter anomalies on DTI. Future research is needed to replicate our findings; candidate genes may also be added to or subtracted from our panel. Similarly to the case with my second dissertation research aim, multivariate approaches that consider imaging maps as a whole (as opposed to treating voxels independently) may improve such predictors of brain integrity. We explored white matter integrity in young, healthy adults. Future research may also find multilocus predictors of white matter structure anomaly patterns specific to particular (stages of) neuropsychiatric disorders. Ideally, such investigation of genes and images together would lead to highly personalized diagnostic, prognostic and therapeutic protocols in the near future.